Received: 29 June 2022 | Accepted: 31 October 2022

DOI: 10.1097/HC9.0000000000000028

ORIGINAL ARTICLE

OPEN

Evaluating sleep in covert encephalopathy with wearable technology: results from the WATCHES study

Adam Buckholz[1] | Lindsay Clarke[2] | Paul Paik[2] | Arun Jesudian[1] | Robert Schwartz[1] | Ana Krieger[3] | Russell Rosenblatt[1] | Robert S. Brown, Jr[1]

[1]Division of Gastroenterology and Hepatology, NewYork-Presbyterian/Weill Cornell Medical Center, New York City, New York, USA

[2]Department of Internal Medicine, NewYork-Presbyterian/Weill Cornell Medical Center, New York City, New York, USA

[3]Department of Medicine, Division of Sleep Medicine, NewYork-Presbyterian/Weill Cornell Medical Center, New York City, New York, USA

Correspondence
Robert S. Brown, Jr, 1305 York Avenue 4th Floor, New York, NY 10065, USA.
E-mail: Rsb2005@med.cornell.edu

Funding information:

This research was funded in part by an NIH T32 Training Grant (Recipient Dr Adam Buckholz). This is an investigator-initiated study with all funding for study materials provided by internal sources at Weill Cornell Medical College.

Abstract

Background and Aims: Covert HE (CHE) is a common early stage of HE associated with poor outcomes. Available neuropsychiatric diagnostic testing is underutilized and has significant clinical limitations. Sleep deterioration is consistently associated with CHE and HE; however, objective data is sparse and it has not been studied longitudinally. We longitudinally study and describe an association of sleep metrics with CHE as detected by a commercial wearable technology.

Methods: We monitored sleep for 6 months using a commercial fitness tracker in 25 participants with cirrhosis, hypothesizing that CHE as diagnosed by psychometric testing would be associated with significant reductions in sleep quality, especially restorative sleep (deep sleep + rapid eye movement). Mixed-effects modeling was performed to evaluate sleep factors associated with CHE and developed and internally validated a score based on these sleep metrics for associated CHE.

Results: Across 2862 nights with 66.3% study adherence, we found that those with CHE had consistently worse sleep, including an average of 1 hour less of nightly restorative sleep, driven primarily by reductions in rapid eye movement. A model including albumin, bilirubin, rapid eye movement, sleep disturbances, and sleep consistency showed good discrimination (area under the receiver operating curve = 0.79) for CHE status with a sensitivity of 76% and specificity of 69%.

Abbreviations: AUROC, area under the receiver operating curve; BMI, body mass index; CTP, Child-Turcotte-Pugh; CLDQ, Chronic Liver Disease Questionnaire; CHE, covert HE; MELD, model for end-stage liver disease; OHE, overt HE; PSQI, Pittsburgh Sleep Quality Index; PSG, polysomnography; PHES, Psychomotor Hepatic Encephalopathy Score; QoL, quality of life; RAPA, Rapid Assessment of Physical Activity; REM, rapid eye movement; WATCHES, WearAble Technology to evaluate the association of Covert Hepatic Encephalopathy with Sleep Study.

Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are provided in the HTML and PDF versions of this article on the journal's website, www.hepcommjournal.com.

Conclusions: Our large longitudinal study of sleep in cirrhosis suggests that sleep derangements in CHE can be detected using wearable technology. Given the known importance of sleep to overall health and CHE/HE to prognosis in cirrhosis, the ability to associate dynamic sleep metrics with CHE may in the future help with the detection and passive monitoring as factors that precipitate decompensation of cirrhosis become better understood and mobile health data validation and integration improves.

INTRODUCTION

Cirrhosis is a leading cause of morbidity and mortality worldwide. HE is a late-stage manifestation of hepatic cirrhosis defined as alterations in neuropsychiatric functioning, and has a broad range of symptoms and severity.[1] Covert HE (CHE) is the earliest and most common manifestation of HE with a prevalence of ~50% in cirrhosis, characterized by changes in motor coordination, concentration, and visuospatial abilities.[2] It is associated with an increased risk of hospitalization and death, as well as reduced quality of life (QoL) and a 20% annual risk of progression to overt HE (OHE).[3] Current options for CHE diagnosis assess neurocognition and include paper-pencil tests such as the Psychomotor Hepatic Encephalopathy Score (PHES), computerized tests such as critical flicker frequency, or validated smartphone-based apps such as EncephalApp (Richmond, VA), which utilizes the Stroop score.[4] While PHES is often considered the gold standard "pen and paper" test,[5] all assessments of CHE are subject to significant confounding, and guidelines recommend combinations of assessments to improve reliability. Despite widespread availability, these assessments are too cumbersome for regular use, with some studies estimating <10% utilization in practice.[6] Limitations to utilization and adherence include the need for specialized training or equipment, time constraints, and lack of definitive management after diagnosis.[1] Similarly, these tools have been rarely evaluated for longitudinal monitoring, rather than initial diagnosis, of CHE. In one study, researchers asked patients to download and use EncephalApp daily to evaluate possible utility in remote monitoring, but adherence was low with <30% of patients demonstrating regular use at even 2 weeks despite training.[7] Additional diagnostic and monitoring tools are needed.

Sleep changes in HE are described most commonly as disturbances in the sleep-wake cycle.[8] The pathophysiology is likely multifactorial, including adenosine-mediated hyperammonemic effects on arousal,[9] reduced ghrelin level,[10] and alterations in endogenous melatonin metabolism in cirrhosis with resultant circadian rhythm dysfunction.[11] Prior research confirms that patients with CHE likewise suffer from suboptimal sleep with associated QoL reduction.[12] Objective sleep data in CHE are sparse but suggest that patients with CHE experience increased sleep latency and fragmentation as well as a reduction in restorative sleep, which consists of time spent in rapid eye movement (REM) and stage 3 (deep) sleep.[13] Despite the known importance of sleep for the maintenance of overall health and the recognition of sleep impairment as a manifestation of HE, there is a paucity of research on dynamic sleep changes in cirrhosis, while small studies evaluating therapeutics such as melatonin or golexanolone rely on subjective measurements of sleep improvement.[14,15]

The gold standard of sleep measurement is monitored polysomnography (PSG), which requires the application of electroencephalography electrodes to the scalp, electromyography electrodes to the limbs, electrooculography, electrocardiography, and pulse oximetry under observation by a trained physician at considerable expense.[16,17] Research-grade actigraphy is more affordable, utilizing accelerometer-based movement detection, but relies on user reporting of sleep time and cannot discriminate sleep stages.[18] These limitations render established methods incapable of real-world sleep tracking, especially in a nonresearch setting. Wrist-worn consumer activity and sleep trackers (hereafter "wearables") leverage accelerometry and photoplethysmography (heart rate detection, heart rate variability) to categorize "2-stage" (sleep vs. awake) and "4-stage" (awake, light, deep, and REM) sleep.[19] Wearables improve on existing monitoring methods in that they do not require the user to input bed and rising times, as integrated photoplethysmography automatically detects changes in autonomic nervous system activity.[20] Likewise, automatic detection capability addresses some of the previously described adherence issues and biases noted in other studies.[21] The WHOOP 3.0 fitness tracker (Whoop Inc.) is a multisensor device incorporating a triaxial accelerometer, optical sensor, capacitive touch sensor, and ambient temperature sensor. It has been validated for 4-stage sleep detection against PSG, and for REM sleep, in particular, it demonstrated adequate accuracy for

identification compared with PSG.[22] Wearables have increased in popularity in the last decade with more than 350 million units shipped in the last calendar year.[23] Their ease of use, affordability, and promising sensor performance make them an attractive target for research.[24] The WearAble Technology to evaluate the association of Covert Hepatic Encephalopathy with Sleep (WATCHES) Study was performed to evaluate the feasibility of noninvasive sleep monitoring to discriminate patients with CHE, and in doing so to assess differences in sleep quality among those with CHE.

METHODS

Study design and participation

Over a 12-month period from 2020 to 2021, patients of the Weill Cornell Medicine Center for Liver Disease and Transplantation were screened and prospectively enrolled. Eligible individuals were adults with a history of Child-Turcotte-Pugh (CTP) Class A or B cirrhosis as diagnosed by a combination of laboratory values, imaging, clinical assessment, and/or pathology. Exclusion criteria included a history of liver transplantation, TIPS procedure, a history of OHE, current/prior use of lactulose or rifaximin, and inability to comprehend and read English, as all study material was in English. To reduce confounding, participants were also excluded if they had active alcohol use >7 drinks weekly, or use of neuromodulatory medications including opioids, prescription sleep aides, and antipsychotics. To avoid confounding with other sleep disorders, those with chronic neuropsychiatric disease, severe chronic obstructive pulmonary disease, obstructive sleep apnea, or body mass index (BMI) >35 were excluded. Eligibility was evaluated using chart review and confirmed by the participant's primary hepatologist in combination with an enrollment interview. Baseline PHES and Stroop testing (with EncephalApp) were performed as well as an assessment by a hepatologist. Those with West Haven grade 0 or 1[25] and PHES ≤ -4 were considered to have CHE, and those with West Haven 0 and PHES > -4 were considered to have no CHE.[26] In cases of disagreement between US-based norms for CHE by EncephalApp and PHES, CHE status was adjudicated by PHES. Other data collected included baseline standard of care laboratory values and a questionnaire including validated subjective assessments of sleep (the Pittsburgh Sleep Quality Index, PSQI), exercise (Rapid Assessment of Physical Activity, RAPA), and liver disease QoL (Chronic Liver Disease Questionnaire, CLDQ), repeated at 3 and 6 months. The study was performed in accordance with Helsinki ethics guidelines and approved by the IRB at Weill Cornell Medical College protocol 20-03021707-05 with informed consent from participants.

Sleep measurement

Participants were managed medically according to the standard of care. Objective sleep data were collected using the WHOOP 3.0 wearable, leased directly from WHOOP Inc. using institutional funds. Participants were provided the tracker free of charge for the duration of 6 months of the study and received a printed and in-application usage tutorial. Anonymity was maintained by generating a dummy name, user ID, and email in the tracker app. Participants were encouraged to wear the tracker continuously and maintain charge (every 5–7 d). Participants wore the tracker on whichever wrist was most comfortable and personal use the application if desired. Anonymous technical support was available. Scheduled follow-up clinic visits were conducted at 3 and 6 months with assessment of HE by West Haven Criteria. At 6 months, repeat laboratory testing was performed and study equipment was returned. As this study took place entirely during the ongoing COVID-19 pandemic, a failure to return in person at 3 or 6 months was not considered study nonadherence; participants could complete follow-up surveys and laboratory testing remotely. As per protocol a maximum of 6 months (first 190 d) of data was analyzed.

Data analysis

The primary outcome of interest in this study was average nightly restorative sleep by CHE status, as defined by REM sleep plus slow wave (or stage 3, "deep") sleep. Data from all participants were analyzed regardless of completion of the full 6-month (190 d) collection. Therefore, data from partial participation was analyzed for the primary outcome as long as it was collected while a participant was eligible and enrolled. Adherence was assessed by the number of days with sleep data available/total days in the study. Missing data from participant voluntary withdrawal was considered a nonadherent night until the full 180-day study period, given that one study aim was to evaluate feasibility of longitudinal monitoring. Missing data from patient withdrawal due to an excluding medical event was not considered nonadherence, nor was withdrawing before data collection.

Analysis of baseline characteristics and individual level sleep data were performed using χ^2 or Wilcoxon rank-sum test for nonparametric means, as appropriate. Factors that were considered possible confounders for sleep alteration were chosen and collected *a priori*, including age, sex, BMI, presence of ascites, and widely accepted scores for liver disease severity (MELD-Na and CTP). Subjective and objective measures of sleep were compared on the individual level and evaluated for concordance using the *F* test of equality for means and variances. Sleep data collected and reported by the

WHOOP 3.0 include time in bed, sleep duration, 4-stage sleep discrimination (awake, light, deep, and REM), sleep consistency, sleep efficiency, heart rate variability, resting heart rate, mean respiratory rate, and number of disturbances. Sleep consistency is a calculated variable which is a rolling 4-day measurement of intraindividual variability in sleep habits measured by bed and wake times and reported on a scale of 0–100%, where 100% means identical sleep and wake times each day. Sleep efficiency refers to the percentage of "in bed" time spent sleeping rather than awake.

Associated sleep factors with CHE

To evaluate for association of sleep factors with CHE status, we performed mixed effects modeling with maximum likelihood estimation, utilizing baseline characteristics (such as sex, albumin) as time invariant factors and participant ID as a random effect independent variable across the model. In essence, this allows that while some factors which impact sleep may be fixed (eg, age and BMI), variation between individuals may be random (due to unmeasured person-level factors) and account for random night-to-night variation. In addition, season-to-season variation was controlled for, as for example the ambient light of a night in July may impact sleep differently than the ambient night in January. Similarly, the data was evaluated over time to note whether the measured individual changes were consistent, which also serves to evaluate a potential bias. Participants are granted access to their sleep data in the WHOOP application, so it is conceivable that noted differences in sleep quality could converge due to the change in participant behavior. Given high levels of noise expected in night-to-night sleep data, a rolling 7-day average was utilized to measure over time reliability of sleep in those with at least 14 days of data collection.

To further evaluate future potential of wearable-based sleep data for identifying CHE, putative discriminatory factors were chosen based on previous studies of sleep in encephalopathy and data from WATCHES. These included age, sex, albumin, bilirubin, presence of ascites, REM sleep duration, sleep disturbances, sleep cycles, total sleep time, sleep efficiency, restorative sleep, slow-wave sleep, and sleep consistency. The association with each of these and presence of CHE was evaluated with logistic calibration, and repetitive iterations were performed to maximize simplicity and discriminative ability as defined by area under the receiver operating curve. The model that was deemed to have the highest potential reproducibility in clinical practice was evaluated for internal validity by bootstrapping with 200 random samples and replacement, with further analysis performed for possible result clustering by individuals.

RESULTS

Participant characteristics

A total of 25 participants enrolled over the study period. Most (68%) of the participants (17/25) were male, and approximately half (12/25) had CHE by PHES. Before initiation of data collection, 3 participants withdrew from the study: one had a significant unrelated medical event, another was lost to follow-up without initiating data collection, and another found the wrist strap too uncomfortable to wear. Another 3 participants were unable to complete the study due to worsening health conditions, dropping out after initiating data collection: one had an OHE episode requiring hospitalization and subsequent rifaximin use after collecting 10 days of data, another received a living donor liver transplantation after 67 days, and another was diagnosed with severe pulmonary hypertension after 67 days requiring long-term neuromodulatory medications, leading to study withdrawal (Figure 1). Besides the one participant who dropped out due to discomfort of the strap, no other participants reported any adverse events related to the study tracker such as skin rash.

In addition to withdrawals due to excluding conditions, one participant stated that he lost the charger and declined a replacement after 5/180 days, and one other stopped collecting data after 9/180 days and declined a trouble-shooting study visit. Because these latter 2 patients remained eligible despite lack of data collection, they were included in adherence calculations. Overall adherence was 66.3%; excluding all study withdrawals, adherence in those who completed the study was 71.3%. Considering the 17 participants who completed the study, the mean number of recorded nights was 158 (minimum 11, maximum 190, median 171, interquartile range 31). A total of 2862 nights of data were analyzed from the 22 participants who collected data, for a mean of 130 nights/participant.

Of those who initiated data collection ($n = 22$), the proportion of women was higher in the CHE group (45.5%) as compared with the non-CHE group (10%). The average MELD-Na and CTP scores were numerically higher without reaching significance in those with CHE (Table 1). Albumin was, on average, slightly lower in those with CHE, and those with CHE were more likely to have ascites or medication for ascites on enrollment. Otherwise, there was no significant difference between groups by baseline characteristics. The mean PHES score among those with CHE was −5.63, while among those without it was −1.27. There was only slight agreement between the PHES test and EncephalApp (Cohen $\kappa = 0.13$) for diagnosing CHE. PHES and EncephalApp agreed on CHE status for 14/25 participants.

Patient-reported outcomes

By the RAPA Physical Activity scale, those with CHE were more likely to report being underactive (81.2% vs.

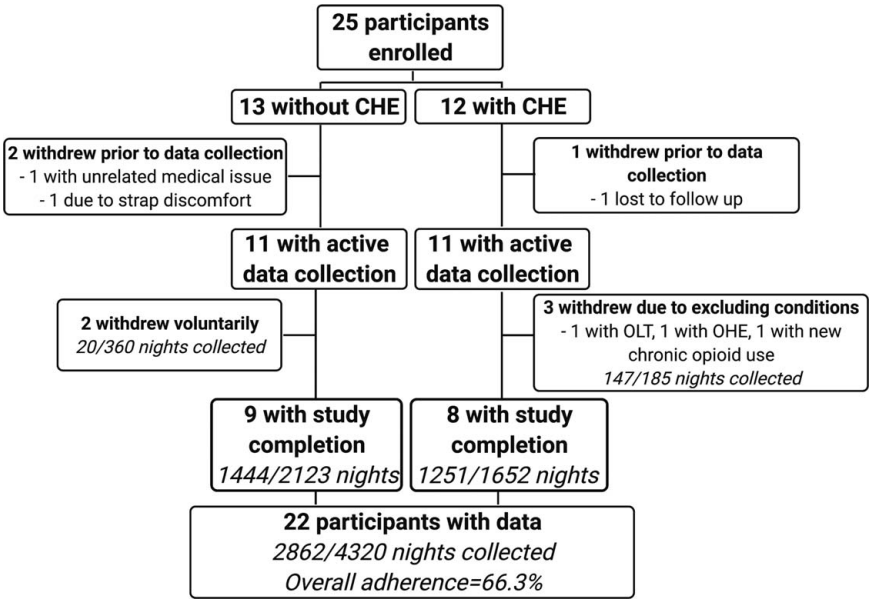

FIGURE 1 CONSORT graph demonstrating participant enrollment by cohort and dropout status with overall study adherence by number of nights collected divided by the number of possible nights. Abbreviations: CHE, covert HE; OHE, overt HE; OLT, refers to orthotopic liver transplant.

36.7%, $p = 0.03$). By this scale, "active" is defined by answering affirmatively to either 3 days weekly of vigorous exercise or 5 days weekly of moderate exercise. There was a trend toward a lower average health-related QoL as assessed by the global CLDQ score in the CHE group as compared with the non-CHE (Supplemental Table 1, http://links.lww.com/HC9/A86), without reaching statistical significance (6.18 vs. 5.74, $p = 0.17$). The global PSQI was marginally elevated in the CHE group as compared with the non-CHE, without achieving statistical significance (5.2 vs. 6.2, $p = 0.46$, although a trend toward a higher rate of sleep disturbances was noted in the CHE group ($p = 0.09$). The global PSQI is consistent with "poor sleep" at a score > 5, and more than half (67%) of the participants across the full cohort reported subjectively poor sleep. Subjective participant estimate of average sleep duration was longer than the actual measured duration by 0.89 hours (Supplemental Figure 1, http://links.lww.com/HC9/A86). This discordance was statistically significant ($p = 0.01$) and similar across the CHE and non-CHE groups.

Association of clinical factors with sleep

In univariable analysis on the individual average level (n = 22), participants with CHE had significantly lower average % of nightly REM (20.8% vs. 15.4%, $p = 0.045$) (Supplemental Table 2, http://links.lww.com/HC9/A86). No baseline characteristic (including age, sex, BMI, cirrhosis etiology, current ascites, MELD, CTP score or years of education) was independently associated with any measured characteristic in univariable analysis, although sex approached significance ($p = 0.06$).

In the mixed effects model, night level data was controlled for the fixed effects of all factors significantly different between the 2 cohorts (sex, albumin, ascites) as well as those with known impacts on sleep (BMI, age, season that sleep night was collected) plus random within effects of individuals. In this model, several of the associations with sleep markers and CHE were attenuated, but multiple factors remained significantly worse in those with CHE (Table 2). This includes our primary outcome of interest, nightly restorative sleep, where those with CHE had 0.95 less hours on average nightly (95% CI: −1.54, −0.36, $p = 0.002$). In addition, those with CHE had less REM sleep (1.42 vs. 1.01 h, $p < 0.001$) and fewer sleep cycles (3.9 vs. 3.1, $p < 0.001$). The reduction in sleep consistency in CHE approached significance (−6.9%, $p = 0.053$).

In separate sensitivity analyses replacing 2 of the components of CTP score (albumin and ascites) with the CTP score, all of these variables remained significant and no additional variables became significant. Similarly in sensitivity analysis controlling only for variables found to be significantly associated in univariable analysis (ascites, sex, albumin), no variables previously found to be significant in mixed effects modeling were found to be insignificant, and no previously insignificant associations became significant. Likewise, in sensitivity analysis including only nights from those who completed the study, all factors remained significant at $p < 0.05$.

Most of the difference in restorative sleep seen between cohorts was a drastic reduction in nightly REM sleep seen in CHE (Figure 2A), where those without CHE had 41% more REM sleep nightly than those without (1.42 vs. 1.01 h). As a function of total nightly sleep in our mixed effects model, those with CHE spent

TABLE 1 Baseline participant characteristics by CHE status

	No CHE (n = 11)	CHE (n = 11)	Total	p[a]
Sex, male, n (%)	10 (90.9)	5 (45.5)	15 (68.2)	0.02
Age (y)	58.4	57.6	58.0	0.51
BMI (kg/m^2)	28.1	27.1	27.6	0.67
Education (total y)	17.2	15.7	16.5	0.21
Cirrhosis etiology				0.69
Alcohol	3	2	5	
Hepatitis B/C	3	5	8	
NAFLD	5	3	8	
other	0	1	1	
MELD-Na	8.7	11.9	10.3	0.25
CTP score	5.8	7.0	6.4	0.09
Sodium (mEq/L)	139.4	138.4	138.9	0.48
Creatinine (mg/dL)	0.89	1.08	0.98	0.17
Total bilirubin (mg/dL)	1.35	2.47	1.91	0.69
Albumin (g/dL)	4.2	3.7	3.9	0.04
Platelet count (u/μL)	90.8	104.9	97.9	0.43
Any ascites, n (%)	2 (18.2%)	9 (81.8%)	11 (50%)	< 0.01

[a]*p* Value determined by Wilcoxon rank-sum test.
Abbreviations: BMI, body mass index; CHE, covert HE; CTP, Child-Turcotte-Pugh; MELD, model for end-stage liver disease.

9.4% less of their total sleep time in REM than those without (Figure 2B), and after controlling for confounders in the model, a sleep night with less than the cohort median 18% was associated with an OR of 11.95 for CHE (95% CI: 8.10–17.62, $p < 0.001$).

Given that both cohorts spent similar time in bed (~7 h nightly) with similar duration of nightly sleep, further graphical analysis was performed to evaluate this finding. A visual inspection of time spent in sleep stages as a function of total time spent in bed demonstrates that all participants increased light sleep periods preferentially to deep or REM sleep when spending more time in bed (Figure 3). The average yield of increased time spent in bed for deep and light sleep was similar in both cohorts, but those with CHE averaged much less REM for each additional hour spent in bed (8.7 vs. 12.6 min, $p < 0.001$) compared with those without.

To analyze whether the noted reduction in mean nightly cycles among those with CHE was responsible for the reduced REM experienced by those with CHE, we examined the relationship between nightly sleep cycles and percentage of sleep spent in REM by CHE status. Across the full cohort, 95.1% of sleep nights consisted of between 1 and 8 cycles (Figure 4), and at each number of cycles the REM sleep achieved is lower in the CHE cohort. When additionally controlling for the number of sleep cycles in mixed effects modeling, those with CHE on average still had a lower REM sleep percentage (−6.05%, $p = 0.002$).

Night-to-night variability in REM sleep was high, but in mixed effects modeling there was moderate individual level consistency (ICC = 0.214). Evaluating 7-day smoothed REM over time graphically demonstrates significant night-to-night variation in each participant and certainly between participants, but when considering the CHE versus non-CHE cohort, REM sleep duration was lower in the CHE cohort throughout the study period (Figure 5).

Discriminating CHE with a sleep score

Secondary analysis was performed to evaluate the performance of sleep metrics in discriminating between those with and without CHE. Raw analysis of REM sleep duration on the individual level demonstrated a sensitivity of 64% (7/11) and specificity of 82% (9/11) for identifying CHE at an average duration of 1.25 hour per participant, with a positive predictive value of 78% (7/9). If excluding those with <14 days to establish a baseline within the device, sensitivity is 70% and specificity 75%.

We found that age, sex, and baseline albumin and bilirubin had the largest impact on discrimination of CHE, while measured REM, sleep disturbances, and sleep consistency also improved discrimination. Age, sleep efficiency, number of sleep cycles, slow-wave sleep, total restorative sleep, and total sleep time did not significantly alter the model performance. While ascites and sex did improve model performance, the subjectivity of ascites and the large sex disparity in our study (only 1 female did not have CHE by PHES) led us to question their potential validity and exclude them from the model. A model consisting of baseline albumin and bilirubin, and average sleep consistency, REM sleep percentage of total sleep and sleep disturbances within an individual had an area under the receiver operating curve for diagnosed CHE by PHES of 0.91 (Figure 6A). The relatively simple model is:

$$\text{CHE sleep score} = 12 - \frac{9.8 \times \text{albumin} + \%\,\text{REM}}{7} + \frac{\text{sleep disturbances} + 2 \times \text{bilirubin} + 0.5 \times \text{sleep consistency}}{5}.$$

Internal validation with bootstrapping on a single night level was performed, which demonstrated

TABLE 2 Association of sleep factors with CHE in mixed effects modeling

	No CHE (n = 1459)	CHE (n = 1397)	ME[a,b] [95% CI]	*p*
Time in bed (h)	6.96	7.02	−0.23 [−1.20, 0.73]	0.64
Nightly sleep (h)	5.86	5.61	−0.56 [−1.58, 0.45]	0.28
Wake time (h)	1.11	1.42	0.33 [−0.16, 0.81]	0.19
Nightly disturbances	8.03	9.38	1.34 [−0.62, 3.39]	0.18
Sleep efficiency (%)	84.4	80.2	−4.75 [−11.5, 1.98]	0.17
Nightly sleep cycles	3.9	3.1	−1.61 [−2.35, −0.86]	< 0.001
Resting heart rate (bpm)	59.7	67.8	1.71 [−8.33, 11.75]	0.74
Heart rate variability	39.3	30.5	−3.71 [−19.62, 12.20]	0.65
Nightly REM sleep %	20.5	14.3	−9.4 [−13.8, 5.0]	< 0.001
Nightly restorative sleep %	36.9	30.3	−12.2 [−19.3, −5.1]	0.001
Sleep consistency %	66.0	57.9	−6.9 [−13.8, 0.10]	0.053
Average REM (h)	1.42	1.01	−0.75 [−1.09, −0.42]	< 0.001
Average deep sleep (h)	1.14	1.16	−0.20 [−0.54 , 0.14]	0.26
Average restorative sleep (h)	2.56	2.16	−0.95 [−1.54, −0.36]	0.002

[a]Mixed effects model controls for fixed effects of age, sex, body mass index, presence of ascites, albumin, season, and participant level random effects.
[b]A coefficient referring to the average change associated with CHE relative to non-CHE.
Abbreviations: CHE, covert HE; ME mixed effects; REM random eye movement sleep.

persistent good discriminatory ability (area under the receiver operating curve = 0.79, 95% CI: 0.77–0.81) and a sensitivity of 76% and specificity of 69% for discriminating baseline CHE at a value > 0 (Figure 6B). Positive (70%) and negative (75%) predictive values were also excellent. Internal validity of this model was estimated via bootstrapping (200 iterations) of individual nights split into 2 random samples. Bias was found to be low (0.002) even when including a cluster-robust provision for individual participants (bias = 0.06). In our cohort, those with CHE had a median 68.6% of sleep nights with CHE sleep score > 0, while those without CHE had a median 15.7% of sleep nights with CHE sleep score > 0.

DISCUSSION

In the longest, most comprehensive study of sleep in cirrhosis, we found consistent and potentially clinically meaningful alterations in sleep physiology among those with CHE. Despite similar demographics and baseline laboratory characteristics, participants with CHE had large reductions in objectively measured nightly

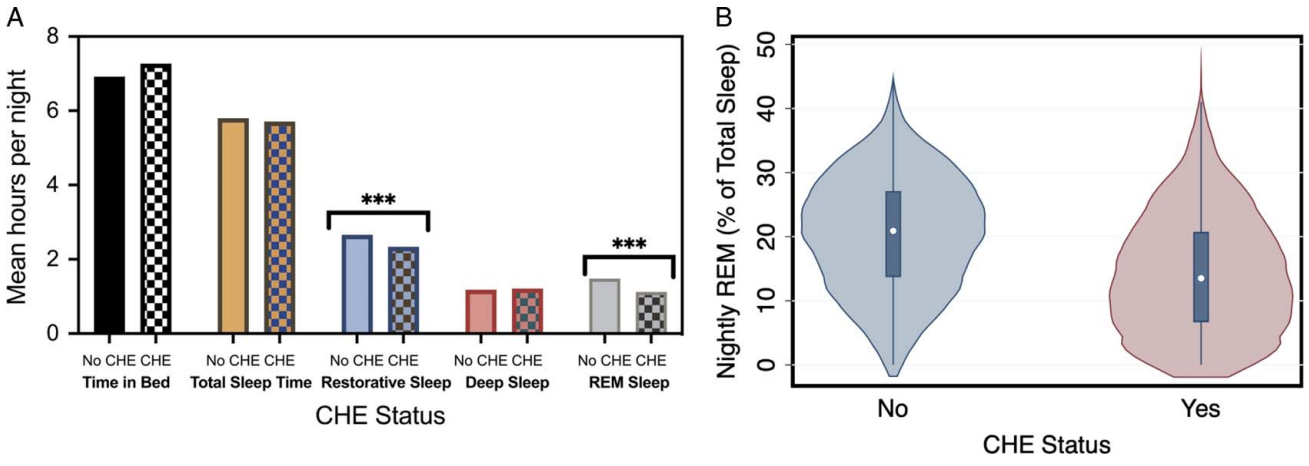

FIGURE 2 Differences in sleep cycle times by CHE status. (A) In mixed effects modeling, those with CHE spent a similar amount of time in bed and had similar total sleep to those without CHE, but with significant (***p <0.001) reductions in restorative sleep, driven by reduction in REM. (B) Violin plot with interquartile range (box) demonstrating significantly lower REM sleep percentage among those with CHE. Abbreviations: CHE, covert HE; REM, rapid eye movement.

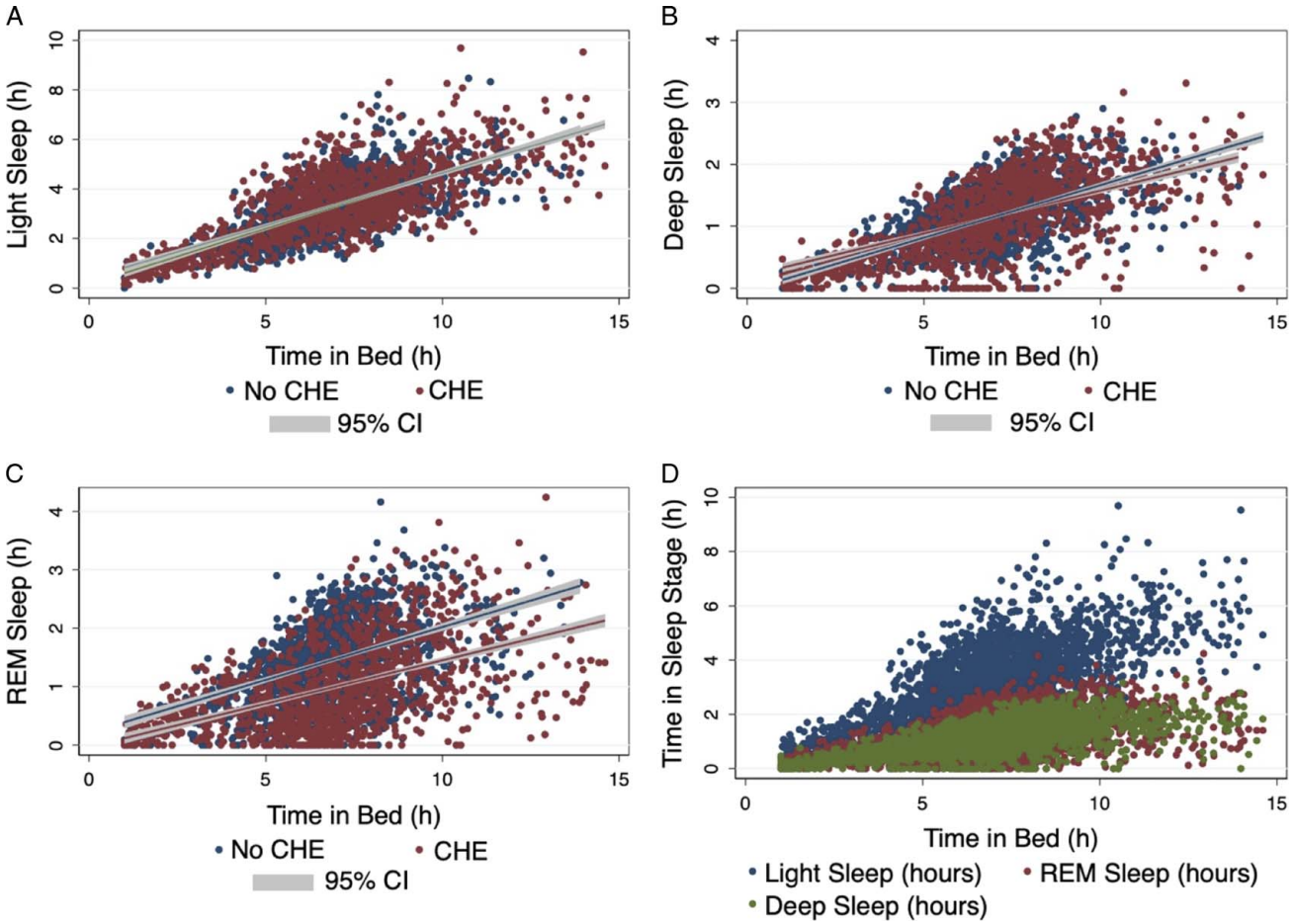

FIGURE 3 Scatter plots of sleep cycle duration over time in bed by cycle and CHE status. Participants with CHE demonstrated similar increases in light (A) and deep (B) sleep with increases in time in bed, while those with CHE did not have the same rate of increase in REM sleep duration, as compared with non-CHE participants (C). In general, all participants increased more light than restorative sleep when spending more time in bed (D). Abbreviations: CHE, covert HE; REM, rapid eye movement.

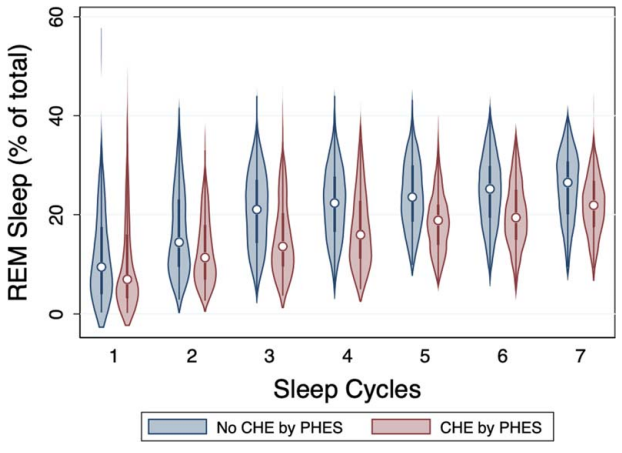

FIGURE 4 Violin plots of percentage of total sleep time spent in REM by the number of sleep cycles and CHE status. Participants with CHE had fewer nightly sleep cycles on average. As the number of measured sleep cycles increased across the study cohort, the percentage of time in REM also increased. However, those with CHE still averaged less REM at each number of nightly cycles. Abbreviations: CHE, covert HE; PHES, Psychomotor Hepatic Encephalopathy Score; REM, rapid eye movement.

restorative sleep, driven primarily by reductions in REM sleep, as measured by a wearable fitness tracker. Likewise, CHE patients showed numerically more sleep disturbances, fewer nightly sleep cycles, and lower night-to-night sleep consistency. Likewise, we internally validated a sleep model with good discrimination for an association with CHE, even on a single night level. While WATCHES was designed as a preliminary study to evaluate the feasibility of remote sleep monitoring in cirrhosis, there are multiple important takeaways from these data.

First, we demonstrated that wearable devices can be used to monitor sleep and detect changes on a nightly level. Since it is known that CHE is characterized by sleep disturbances, the use of wearable technology to detect these subtle changes in sleep in individual patients demonstrates enormous potential. Cirrhosis and HE follow a chronic undulating course punctuated by significant life-threatening events such as progression to OHE. Previously validated diagnostic tools offer only single data points, which have obvious limitations.

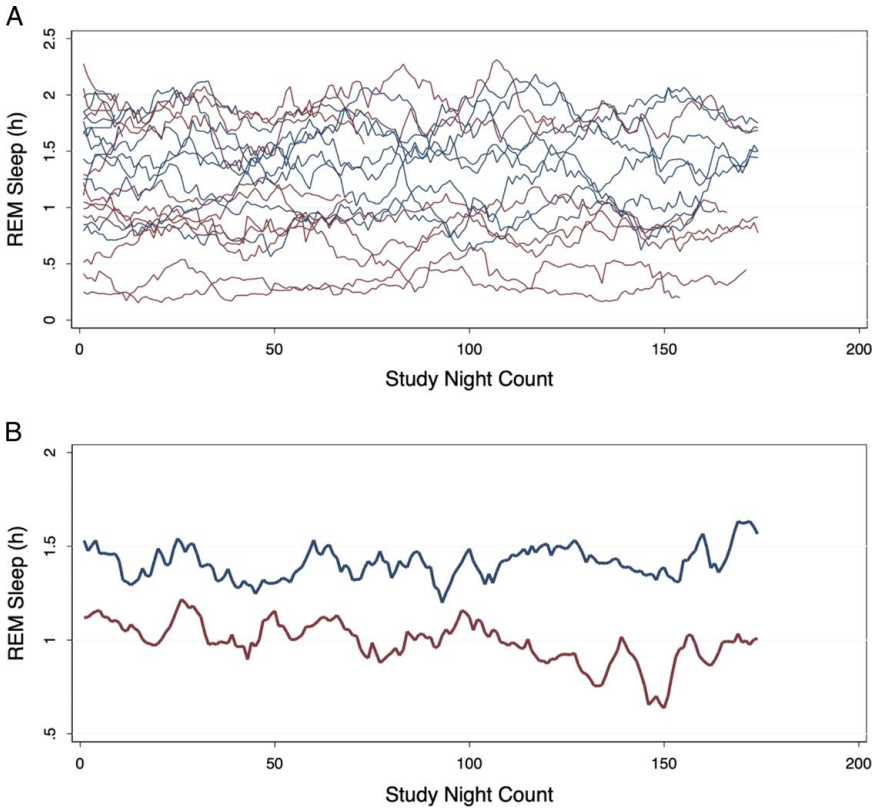

FIGURE 5 Seven day end-to-end rolling average REM over the study period. (A) On the individual level there is considerable variability but those with (red) and without (blue) CHE cluster to the lower and upper REM durations. (B) Over the full study period the difference in average nightly REM for the CHE (red) and non-CHE (blue) cohorts remained significant. Abbreviations: CHE, covert HE; REM, rapid eye movement.

For one, the lack of concordance found in our study for identifying CHE by established methods such as PHES and EncephalApp has been previously noted, and this discordance is exacerbated by a single point of contact.[27] In addition, most OHE events occur in the setting of a specific stressor such as infection, so remote monitoring is an attractive potential tool for real-time event detection and early intervention. Likewise, the relatively high ~70% adherence rate and modest dropout rate exceed that found in the limited prior research with other tools. One potential factor contributing to the improved adherence in WATCHES is that the data (sleep) had obvious real-world relevance to patients, boosting engagement and interest. In addition,

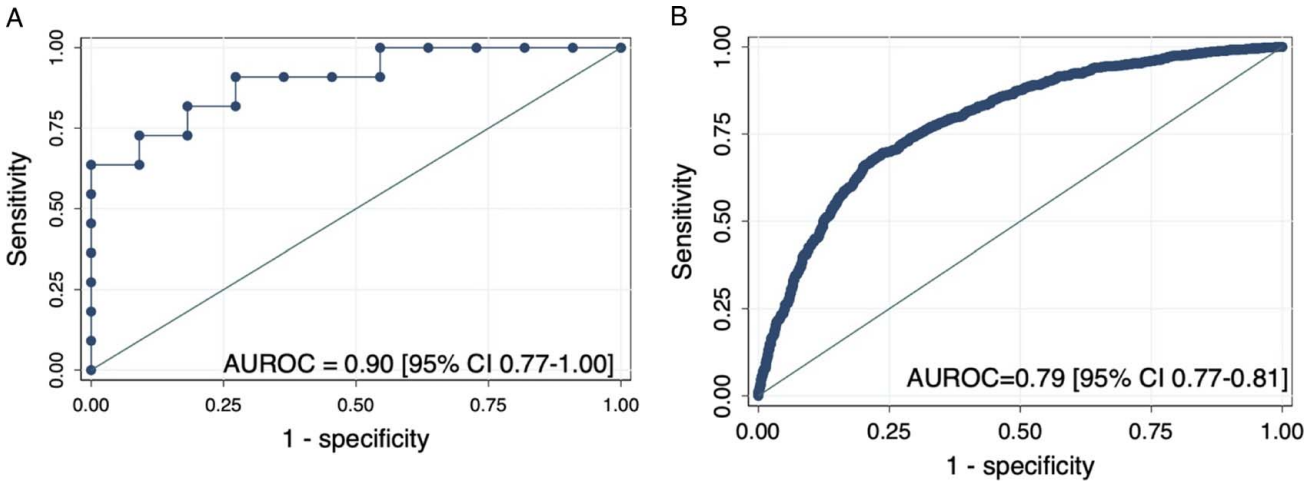

FIGURE 6 Discrimination performance of CHE sleep score as AUROC. (A) The model was derived on the individual participant level (n = 22) using bilirubin, albumin, sleep consistency, sleep disturbances, and rapid eye movement duration. (B) The model was internally validated on the individual night level using bootstrapping with random samples, maintaining good reliability. Abbreviation: AUROC, area under receiver operating curve; CHE, covert HE.

the user interface of consumer wearables, being designed to increase customer engagement with a commercial product, may improve participant understanding, engagement, and retention. While only in the pilot phase, this technology and others like it may become more and more common in clinical medicine, and the early results for retention and feasibility for data collection are promising.

Second, the sleep findings in WATCHES support and expand the findings in smaller studies utilizing PSG, suggesting that previously described single night sleep derangements in CHE are persistent and consistent over long periods of time. While our study was not powered to detect a difference in subjective sleep assessment, the PSQI was insufficient to differentiate the CHE state using qualitative measurement, as noted by the lack of a statistical difference in qualitative sleep quality. Yet, our study was consistent with prior literature evaluating PSQI in cirrhosis in demonstrating that most patients with cirrhosis report subjectively disturbed sleep.[28] In another study evaluating QoL in cirrhosis, researchers found reduced sleep quality, as measured by PSQI, in the majority of participants, and that this reduced subjective quality was associated with decreased health-related QoL.[29] There is a strong pathophysiologic basis to suggest that the differences detected by remote monitoring are clinically relevant in the disease course of HE. Progressive hyperammonemia has been noted to cause sleep architecture changes like those found in our study, including loss of REM sleep.[9,30] Similarly, hyperammonemia alters brainstem dopamine receptors, which are important for transitions into REM sleep.[31] Finally, reduced hepatic melatonin clearance leads to circadian rhythm disruption in cirrhosis, potentially contributing to sleep fragmentation as seen in our study.[32] Fracturing of sleep with fewer nightly cycles may lead to less time spent in REM, which is generally the final sleep stage in a cycle. REM sleep is crucial for memory formation and concentration; progressive loss of REM sleep due to multiple physiologic factors may in fact be causative in addition to emblematic of progressive HE.

Third, our CHE sleep score suggests that measured objective sleep markers may have the potential to serve as an effective diagnostic tool for clinicians' identification of CHE. With the continued improvement of wearable technology, these findings may offer a new path forward for HE, a critically important manifestation of cirrhosis at large. While further validation is needed, utilizing easily measurable sleep metrics in combination with basic patient characteristics diagnosed CHE with high fidelity over >2500 nights even when accounting for individual level variability in mixed effects modeling. Further studies can help elucidate the minimum sleep monitoring time needed to establish a "baseline," and how quickly and consistently clinical deterioration manifests as sleep alterations. Given that those without CHE averaged only 1 in 7 sleep nights with a positive CHE sleep score and those with CHE averaged 2 of 3 nights with a positive sleep score, CHE may be associated with sleep with high fidelity even over a short time.

There are several important limitations to this study. Most notably, data was collected using a commercial wearable, the WHOOP 3.0. Despite multiple validation studies and meta-analyses suggesting that wearables have good performance for sleep stage detection against PSG, this is the first study evaluating this technology in cirrhosis.[22,33] Future research should confirm that wearables, and the WHOOP in particular, maintain similar fidelity against PSG in this population. WHOOP was chosen due to the fact that it has been studied and validated for automatic-mode sleep detection. The use of a single device, however, limits generalizability of these findings, and broad replication in other wearables is needed before any widespread clinical utilization. While this does limit external validity, our study has the advantage of isolating as closely as is feasible a patient population without other potential confounders to sleep. Furthermore, most validation studies for wearables in sleep have been conducted in healthy volunteers. This serves as a limitation in all studies utilizing wearables in a disease state; one previously described paradox of wearables is that those most likely to seek to use them are those least likely to benefit from them.[34] The veracity of these findings is helped, in no small part, by the fact that they are quite consistent with sleep architecture findings from smaller studies based on PSG.[35] As wearables continue to improve, it is imperative to identify whether and how measured sleep metrics from any one wearable can be generalized to the diagnosis and management of disease at large.

The use of commercial technology serves as both an advantage and limitation in this study. While wearables have widespread consumer acceptance, they have only recently been started to gain FDA clearance for clinical indications. This means that despite their availability, they are unlikely to be reimbursed by commercial insurance in the near future, making remote monitoring potentially more costly to the patient than existing diagnostic tools. Given the out-of-pocket cost needed for clinical use in the short term, there are valid questions of equitable access across the spectrum of disease; regardless, rapid consumer uptake suggests that wearables are likely to have persistent and deep penetration across all patient groups. In addition, wearables raise questions of confidentiality when utilized for clinical purposes. In our study, utilizing dummy identifiers allowed for the maintenance of patient privacy, but in the case of widespread clinical use methods will need to be developed to secure identifiable health information. Survey responses suggest participants largely enjoyed and were able to use the WHOOP without significant technical difficulty, but technical savviness may be highly variable in cirrhosis.

The study design also somewhat limits more generalized interpretations. One weakness was the use of imperfect markers for CHE rather than "hard" end points such as overt episodes, hospitalizations, or transplant. As a pilot study, we utilized stringent exclusion criteria to mitigate systematic bias, but further study will need to confirm whether such findings are also seen in those with a history of OHE, active alcohol use, or potentially sleep-modifying medication use, which are common in the cirrhosis population. In a broader population, use of such "hard" clinical end points may be more feasible. This study was designed to determine whether sleep can reasonably discriminate within at-risk populations, and both PHES and Stroop have already been validated for long-term risk of clinical outcomes. Statistically, our study also has limitations. Despite consistent reduction in sleep markers over the duration of the study, night-to-night variability remains high, and our conclusions are based on data collection from only 22 individuals, increasing the opportunity for bias. The use of a random-effects model minimizes this impact, but the sex discrepancy is only one example of how a small sample size may limit applicability. While both PHES and Stroop have adjustments for gender when diagnosing likely CHE, at least one previous study has suggested that women with cirrhosis have worse subjective sleep by PSQI, reinforcing that more broad patient sampling is needed before any sleep data can be used for clinical management of CHE.[12]

Despite these limitations, this study has clear strengths in that it utilizes noninvasive, passive data collection to evaluate a marker with practical importance to patients. Unlike PHES or critical flicker frequency, it does not require specialized training or medical equipment or take substantial time during clinic visits. Unlike other smartphone-based testings, such as EncephalApp, it does not rely on significant patient initiative. In the correct patient and pending further research, sleep monitoring with wearables may offer a substantial leap forward in the pragmatic management of CHE and perhaps HE at large.

AUTHOR CONTRIBUTIONS

Dr Adam Buckholz: conceptualization, methodology, data collection, data analysis, writing (original draft and review). Dr Lindsay Clarke and Dr Paul Paik: methodology, data collection, writing (review and editing). Dr Arun Jesudian: conceptualization, methodology, writing (review and editing). Dr Robert Schwartz: methodology, writing (review and editing). Dr Ana Krieger: methodology, data analysis, visualization, writing (review and editing). Dr Russell Rosenblatt: conceptualization, methodology, data analysis, writing (review and editing). Dr Robert S. Brown: methodology, data analysis, writing (original draft and review), supervision.

CONFLICT OF INTEREST

Arun Jesudian consults for and is on the speakers' bureau for Salix Pharmaceuticals. The study was initiated by the investigators and the wearable fitness trackers were leased from WHOOP Inc. at a predetermined "research price," ~30% lower than market cost per month. The remaining authors declare no conflicts of interest.

ORCID

Adam Buckholz https://orcid.org/0000–0002–0068–353X
Lindsay Clarke https://orcid.org/0000–0002–6744–084X
Paul Paik https://orcid.org/0000–0002–9464–3984
Arun Jesudian https://orcid.org/0000–0002–8562–3375
Robert Schwartz https://orcid.org/0000–0002–5417–5995
Ana Krieger https://orcid.org/0000–0001–9458–0111
Russell Rosenblatt https://orcid.org/0000–0003–3981–7053
Robert S. Brown, Jr https://orcid.org/0000–0002–8220–065X

REFERENCES

1. Vilstrup H, Amodio P, Bajaj J, Cordoba J, Ferenci P, Mullen KD. Hepatic encephalopathy in chronic liver disease: 2014 Practice Guideline by the American Association for the Study of Liver Diseases and the European Association for the Study of the Liver. Hepatology. 2014;60:715–35.
2. Das A, Dhiman RK, Saraswat VA, Verma M, Naik SR. Prevalence and natural history of subclinical hepatic encephalopathy in cirrhosis. J Gastroenterol Hepatol. 2001;16:531–535.
3. Arguedas MR, DeLawrence TG, McGuire BM. Influence of hepatic encephalopathy on health-related quality of life in patients with cirrhosis. Dig Dis Sci. 2003;48:1622–6.
4. Bajaj JS, Heuman DM, Sterling RK, Sanya AJ, Siddiqui M, Matherly S, et al. Validation of EncephalApp, smartphone-based Stroop test, for the diagnosis of covert hepatic encephalopathy. Clin Gastroenterol Hepatol. 2015;13:1828.e1.
5. Weissenborn K. Hepatic encephalopathy: definition, clinical grading and diagnostic principles. Drugs. 2019;79:5–9.
6. Lauridsen MM, Bajaj JS. Hepatic encephalopathy treatment and its effect on driving abilities: a continental divide. J Hepatol. 2015;63:287–8.
7. Louissant J, Lok A, Fortune BTE. Acceptance and use of a smartphone application in cirrhosis. Liver Int. 2020;40:1556–63.
8. Sherlock S, Summerskill WHJ, White LP, Phear EA. Portal-systemic encephalopathy neurological complications of liver disease. Lancet. 1954;264:453–7.
9. Bersagliere A, Raduazzo ID, Nardi M, Schiff S, Gatta A, Amodio P, et al. Induced hyperammonemia may compromise the ability to generate restful sleep in patients with cirrhosis. Hepatology. 2012;55:869–78.
10. Bajaj JS, Saeian K, Schubert CM, Franco R, Franco J, Heuman DM. Disruption of sleep architecture in minimal hepatic encephalopathy and ghrelin secretion. Aliment Pharmacol Ther. 2011;34:103–5.
11. Steindl PE, Finn B, Bendok B, Rothke S, Zee PC, Blei AT. Disruption of the diurnal rhythm of plasma melatonin in cirrhosis. Ann Intern Med. 1995;123:274–7.
12. Labenz C, Baron JS, Toenges G, Schattenberg JM, Nagel M, Sprinzl FM, et al. Prospective evaluation of the impact of covert hepatic encephalopathy on quality of life and sleep in cirrhotic patients. Aliment Pharmacol Ther. 2018;48:313–21.

13. Montagnese S, Middleton B, Mani AR, Skene DJ, Morgan MY. Sleep and circadian abnormalities in patients with cirrhosis: features of delayed sleep phase syndrome? Metab Brain Dis. 2009;24:427–39.
14. Zhao X, Wong P. Managing sleep disturbances in cirrhosis. Scientifica (Cairo). 2016;2016. doi:10.1155/2016/6576812
15. Montagnese S, Lauridsen M, Vilstrup H, Zarantonello L, Lakner G, Fitilev S, et al. A pilot study of golexanolone, a new GABA-A receptor-modulating steroid antagonist, in patients with covert hepatic encephalopathy. J Hepatol. 2021;75:98–107.
16. Malhotra A, Younes M, Kuna ST, Benca R, Kushida CA, Walsh J, et al. Performance of an automated polysomnography scoring system versus computer-assisted manual scoring. Sleep. 2013; 36:573–82.
17. Geessinck FAJ, Pleijhuis RG, Mentink RJ, van der Palen J, Koffijberg H. Cost-effectiveness analysis of the DiagnOSAS Screening tool compared with polysomnography diagnosis in dutch primary care. J Clin Sleep Med. 2018;14:1005.
18. Coma-Del-Corral MJ, Alonso-Álvarez ML, Allende M, Cordero J, Ordax E, Masa F, et al. Reliability of telemedicine in the diagnosis and treatment of sleep apnea syndrome. Telemed J E Health. 2013;19:7–12.
19. de Zambotti M, Cellini N, Goldstone A, Colrain IM, Baker FC. Wearable sleep technology in clinical and research settings. Med Sci Sports Exerc. 2019;51:1538–57.
20. de Zambotti M, Rosas L, Colrain IM, Baker FC. The sleep of the ring: comparison of the ŌURA sleep tracker against polysomnography. Behav Sleep Med. 2019;17:124–36.
21. Stone AA, Shiffman S, Schwartz JE, Broderick JE, Hufford MR. Patient non-compliance with paper diaries. BMJ. 2002;324: 1193–4.
22. Miller DJ, Roach GD, Lastella M, Scanlan A, Bellenger C, Halson S, et al. A validation study of a commercial wearable device to automatically detect and estimate sleep. Biosensors. 2021;11: 185.
23. Depner CM, Cheng PC, Devine JK, Khosla S, de Zambotti M, Robillard R, et al. Wearable technologies for developing sleep and circadian biomarkers: a summary of workshop discussions. Sleep. 2020;43:1–13.
24. Chinoy ED, Cuellar JA, Huwa KE, Jameson J, Watson C, Bessman S, et al. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep. 2021; 44:1–16.
25. Conn HO, Leevy CM, Vlahcevic ZR, Rodgers JB, Maddrey WC, Seeff L, et al. Comparison of lactulose and neomycin in the treatment of chronic portal-systemic encephalopathy: a double blind controlled trial. Gastroenterology. 1977;72:573–83.
26. Ferenci P, Lockwood A, Mullen K, Tarter R, Weissenborn K, Blei AT. Hepatic encephalopathy—definition, nomenclature, diagnosis, and quantification: final report of the working party at the 11th World Congresses of Gastroenterology, Vienna, 1998. Hepatology. 2002;35:716–21.
27. Duarte-Rojo A, Allampati S, Thacker LR, Flud CR, Patidar K, White M, et al. Diagnosis of covert hepatic encephalopathy: a multi-center study testing the utility of single versus combined testing. Metab Brain Dis. 2019;34:289.
28. Ghabril M, Jackson M, Gotur R, Weber R, Orman E, Vuppalanchi R, et al. Most individuals with advanced cirrhosis have sleep disturbances, which are associated with poor quality of life. Clin Gastroenterol Hepatol. 2017;15:1271.
29. Tapper EB, Baki J, Parikh ND, Lok AS. Frailty, psychoactive medications, and cognitive dysfunction are associated with poor patient-reported outcomes in cirrhosis. Hepatology. 2019;69: 1676–85.
30. Llansola M, Cantero JL, Hita-Yañez E, Mirones-Maldonado MJ, Piedrafita B, Ahabrach H, et al. Progressive reduction of sleep time and quality in rats with hepatic encephalopathy caused by portacaval shunts. Neuroscience. 2012;201:199–208.
31. Ciećko-Michalska I, Szczepanek M, Słowik A, MacH T. Pathogenesis of hepatic encephalopathy. Gastroenterol Res Pract. 2012:1–7.
32. Montagnese S, Middleton B, Mani AR, Skene DJ, Morgan MY. On the origin and the consequences of circadian abnormalities in patients with cirrhosis. Am J Gastroenterol. 2010;105:1773–81.
33. Miller DJ, Sargent C, Roach GD. A validation of six wearable devices for estimating sleep, heart rate and heart rate variability in healthy adults. Sensors (Basel). 2022;22:1–17.
34. Smuck M, Odonkor CA, Wilt JK, Schmidt N, Swiernik MA. The emerging clinical role of wearables: factors for successful implementation in healthcare. NPJ Digit Med. 2021;4:1–8.
35. Formentin C, Garrido M, Montagnese S. Assessment and management of sleep disturbance in cirrhosis. Curr Hepatol Rep. 2018;17:52.

How to cite this article: Buckholz A, Clarke L, Paik P, Jesudian A, Schwartz R, Krieger A, et al. Evaluating sleep in covert encephalopathy with wearable technology: results from the WATCHES study. Hepatol Commun. 2023;7: e0002. https://doi.org/10.1097/HC9.0000000000000028