# The diagnostic value of sleep and vigilance tests in central disorders of hypersomnolence

Johannes Mathis[1,2,*,†], Daniel Andres[1,†], Wolfgang Schmitt[3], Claudio L. Bassetti[1], Christian W. Hess[1], and David R. Schreier[1]

[†]**These authors contributed equally to this work.**

1 Department of Neurology, Inselspital, Bern University Hospital, University of Bern, Switzerland

2 Sleep Medicine, Neurozentrum Bern, Switzerland

3 University Hospital of Psychiatry and Psychotherapy, Bern, Switzerland

* Correspondence to: Prof. Dr. med. Johannes Mathis
Department of Neurology, Inselspital, Bern University Hospital, University of Bern,
CH-3010 Bern, Switzerland
Phone: +41 31 630 60 70
Email: johannes.mathis@belponline.ch

ORCID numbers:

J. Mathis: https://orcid.org/0000-0002-3422-2715
D. Andres: https://orcid.org/0000-0001-6301-1784
W. Schmitt: https://orcid.org/0000-0002-5587-6240
C. L. Bassetti: https://orcid.org/0000-0002-4535-0245
D. R. Schreier: https://orcid.org/0000-0001-6088-2015

# Abstract

**Study Objectives**: This retrospective cross sectional observational study explored the diagnostic value of selected sleep and vigilance tests (SVT) beyond the multiple sleep latency test to differentiate between various central disorders of hypersomnolence (CDH) and fatigue syndromes.

**Methods**: Data from patients who underwent the multiple sleep latency test and at least one additional SVT were extracted from the Bern sleep database (1997-2018). 1,352 patients with a CDH (106 narcolepsy type 1, 90 narcolepsy type 2, 119 idiopathic hypersomnia, 192 nonorganic hypersomnia, 205 insufficient sleep syndrome), fatigue syndromes (n=183), and a subgroup of patients with sleep apnoea (n=457) were analysed. Classification based on SVT parameters was compared with the final clinical diagnosis serving as a reference.

**Results**: An overall model predicted the final diagnosis in 49.5% of patients. However, for the pairwise differentiation of two clinically suspected diagnoses, many SVT parameters showed a sensitivity and specificity above 70%. While the overall discrimination power of the multiple sleep latency test was slightly better than the one of the maintenance of wakefulness test, the latter differentiated best between narcolepsy and idiopathic hypersomnia with prolonged sleep need. Disproportionally poor results in reaction tests (e.g. steer clear test), despite comparable or lower sleepiness levels (SLAT, WLAT), were valuable for differentiating nonorganic hypersomnia from idiopathic hypersomnia/sleep insufficiency syndrome.

**Conclusion**: This study demonstrates how the combination of a careful clinical assessment and a selection of SVTs can improve the differentiation of CDH, whereas it was not possible to establish an overall prediction model based on SVTs alone.

**Keywords:** narcolepsy; idiopathic hypersomnia; nonorganic hypersomnia; maintenance of wakefulness test; multiple sleep latency test, vigilance tests, hypersomnolence

**Statement of significance**

The differential diagnosis of central disorders of hypersomnolence is challenging but of great importance for patient care. Unlike the multiple sleep latency test, vigilance tests have not been routinely applied so far for diagnostic purposes. This is the first large-scale retrospective cross-sectional observational study analysing the diagnostic value of many different sleep and vigilance tests. The results suggest that selected vigilance tests can improve the identification of the final diagnosis out of the clinically determined suspected diagnoses. Particularly for the difficult differentiation of narcolepsy without cataplexy and idiopathic hypersomnia with prolonged sleep need, the maintenance of wakefulness test seems to be of great additive value. Furthermore, performance tests help to separate nonorganic hypersomnia from organic disorders of hypersomnolence.

# Introduction

"Hypersomnolence" was introduced in 2014 as an umbrella term by the American Association of Sleep Medicine in the chapter title "Central disorders of Hypersomnolence" of the most recent International Classification of Sleep Disorders (ICSD-3) [1]. The chapter covers a broad spectrum of disorders whose main symptoms range from excessive daytime sleepiness (EDS), hypersomnia, to tiredness, and fatigue. Therefore, in this study the term "hypersomnolence" will be used as an umbrella term as well, covering all of the above symptoms [1]. In an attempt to disentangle these subjective complaints as accurately as possible, physicians are faced with the fact that all these characteristics can overlap and/or appear simultaneously in the same patient [2]. As a result, these terms are often confused or used interchangeably by the patients and sometimes also by physicians or researchers. However, differentiating its precise meaning, as described below, will help to establish the correct diagnosis.

EDS is the cardinal symptom in many central disorders of hypersomnolence (CDH) and the ICSD-3 defines EDS as the inability to maintain wakefulness and alertness during the major waking episodes of the day, with sleep occurring unintentionally or at inappropriate times almost daily for at least three months [1]. Sleepiness is mainly perceived in passive situations and can be masked at least temporarily by any type of activity. The prevalence of EDS in a healthy population is estimated between two to 28%, depending on methodological approaches, definitions, and populations under investigation [2, 3]. EDS is often mistaken for a disease in itself, but it should rather be understood as a severe and disabling symptom of many different underlying disorders [4].

The term 'hypersomnia' is generally used in association with specific disorders of hypersomnolence, such as idiopathic hypersomnia or periodic hypersomnia in Kleine-Levin Syndrome characterised by prolonged sleeping periods of 11-14 hours, a feature which was recently called "excessive need for sleep", or "pleiosomnia" [5, 6, 7, 8]. However, a precise definition of 'hypersomnia' on a symptom level does not exist yet. Patients with hypersomnia experience EDS despite prolonged sleeping periods which is in contrast to 'long sleepers' (normal variant) who also need more sleep (often >10 hours) but do not experience EDS when sufficiently rested [1]. Typically, patients with hypersomnia report prolonged difficulties of waking up in the morning (sleep inertia) and frequent long and unrefreshing naps.

'Tiredness' is a poorly defined term. In healthy adults, tiredness usually implies a reversible state of physical or mental exhaustion. In daily activities, tiredness can be understood as a biological consequence of decreased reserve capacities and decreased resistance to stressors which can be restored through rest but not necessarily requiring sleep. It is often used interchangeably with lack of energy and/or lack of initiative and typically occurs in patients with depressive symptoms [9].

'Fatigue' is another widely used term throughout literature which lacks a coherent definition. In relation to fitness to drive, the term is often used synonymously with EDS [10, 11]. In a clinical context, fatigue is one of the most disabling symptoms, and its meaning is closer to that of tiredness [12]. Fatigue was originally defined as the decline in performance, physical or mental, carried out over time ('time on task performance decrement') [13, 14]. With increasing fatigue, it becomes more difficult to carry out a certain task, and therefore more psychic energy is needed to maintain

performance. Consequently, fatigue may be defined as a subjectively experienced aversion to investing further effort into a given task. It is important to acknowledge that EDS or tiredness may exacerbate fatigue and thereby lead to performance decrements as well [15]. Unlike patients with sleepiness, those with severe fatigue cannot counteract or mask their symptoms by physical or mental activity, as this rather increases their exhaustion [15].

In summary, CDH include a broad spectrum of disorders which are all difficult to separate from each other and also from (chronic) fatigue syndromes (FS) by clinical means [1]. The CDH-spectrum ranges from narcolepsy with (NT1) and without cataplexy (NT2), idiopathic hypersomnia, periodic hypersomnia (Kleine-Levin-Syndrome) and insufficient sleep syndrome, nonorganic hypersomnia (hypersomnia associated with a psychiatric disorder), to hypersomnia due to medical disorders or medication/substance.

It is not the aim of this paper to repeat the well-established clinical and paraclinical characteristics, definitions and the pathomechanisms of the CDH disorders, which are precisely described in the ICSD manuals [1].

Some of these disorders, such as NT1, hypersomnia due to medical disorders or due to substances, are easy to diagnose based on their pathognomonic clinical presentation or by specific laboratory findings. However, most CDH are rather difficult to differentiate clinically and require a battery of sleep and vigilance tests (SVT), complementing the careful long-term history of nocturnal and daytime symptoms [16]. In order to quantify subjective sleepiness, simple questionnaires such as the Epworth Sleepiness Scale (ESS) are used [17], and fatigue or tiredness are assessed by using the Fatigue Severity Scale. Objective measures of sleepiness are obtained by using the multiple sleep latency test (MSLT), to assess sleep propensity, and/or the maintenance of wakefulness test (MWT) to measure the ability to stay awake [18]. A polysomnography (PSG) or video-polysomnography is indicated to prove or disprove any causes of non-restorative sleep, such as sleep-related breathing disorders or sleep-related movement disorders as well as parasomnias or sleep-related seizures. Sleep efficiency in the PSG (PSE) is a valuable biomarker of sleep quality. Wrist actigraphy helps to disclose abnormal sleep-wake rhythms and poor sleep hygiene or long-lasting inactivity periods during the day [16]. In both research and in clinical sleep-wake medicine, various types of more or less sophisticated driving simulators or other vigilance tests such as the Steer Clear Test (SCT), the psychomotor vigilance test (PVT), and the pupillary unrest index (PUI) are used to judge fitness to drive. This is particularly important if a patient's professional activity involves participation in motor vehicle traffic [19, 20].

To the best of our knowledge, these tests have not been used specifically to determine the underlying cause of hypersomnolence so far, apart from the sleep latency in the MSLT (SLAT) and sleep onset REM periods (SOREMPs), but they may yield useful information when differentiating between CDH groups, including differentiation from the fatigue syndromes [16].

Therefore, we aimed to investigate the diagnostic value of frequently collected SVT parameters in the differentiation of CDH beyond SLAT and SOREMPs in the MSLT [20].

The primary aim was to explore possible specific SVT patterns within CDH, known to present great differential diagnostic ambiguities. The first objective was to conduct an explorative data analysis of SVT results collected from patients diagnosed with CDH, fatigue syndromes, or sleep apnoea. The

second objective was to assess the differences in SVT results between the diagnostic groups and to describe their distribution range and possible grouping clusters.

The secondary aim was to assess to which extent SVT parameters could be useful for the differentiation of the diagnoses of interest, particularly for CDH. Hence, the first objective of the secondary aim was to evaluate the diagnostic value of each SVT for each pairwise comparison. The second objective of the secondary aim was to assess whether the combination of multiple SVT parameters could improve the discrimination between multiple ambiguous diagnoses among CDH.

# Methods

## Ethical approval

Ethical approval for the analysis of clinical data was granted by the institutional board of the Inselspital and the local ethical committee (KEK Nr. 185/06, Inselspital-Nr. 1267). Due to the retrospective nature of this study, no protocol was registered.

## Study population and diagnostic system

The data for this retrospective cross sectional observational study were extracted from the Bern Sleep Database, established in 1997 and continuously expanded since [21]. The clinical database contains more than 20,000 SVT results, obtained from all patients with sleep-wake disorders referred to the Sleep-Wake-Epilepsy Centre since 1997. The encrypted database extract consists of 1) demographic and clinical variables (e.g. age, weight), diagnosis and limited information on the medical history (medication); 2) the most important paraclinical parameters, i.e. SVT results, derived from PSG, MSLT, MWT, actigraphy, PVT, SCT, and pupillography; 3) scales and scores (e.g. ESS, Karolinska Sleepiness Scale, Fatigue Severity Score). Patients only underwent the clinically relevant tests, i.e. only few patients underwent all tests.

PSG, MSLT, and MWT were recorded with Somnologica, EMBLA, ResMed, San Diego, Calif., USA. The SCT was performed according to Findley et al. [22]. The PVT was performed with a device from Ambulatory Monitoring Inc. For the calculation of the PUI, a pupillography device from AMTech GmbH, Germany, was used. Actigraphy was recorded with an actimeter from Ambulatory Monitoring Inc. All recordings were performed according to international guidelines [20].

In principle, patients were diagnosed according to the most recent edition of the ICSD, namely ICSD-1 up to 2005, ICSD-2 up to 2014 and ICSD-3 since [1, 23, 24]. The only deviations from the ICSD diagnostic criteria were made for the diagnosis of idiopathic hypersomnia. Since the founding of the sleep centre in Bern, idiopathic hypersomnia was diagnosed exclusively in the presence of prolonged sleep need. Therefore, 'idiopathic hypersomnia' in this study always refers to the subgroup with prolonged sleep need. In the absence of prolonged sleep need, patients complaining exclusively about EDS were diagnosed as 'EDS of unknown origin' a group similar to the ICSD-2 diagnosis

"Physiological (organic) Hypersomnia, unspecified" [23]. This group also included other poorly defined disorders such as "Subjective DOES complaint without objective findings" (DOES: disorders of excessive sleepiness) and "Psychophysiological DOES" according to the first edition of ICSD, in which the category of idiopathic hypersomnia without prolonged sleep did not formally exist [24]. The diagnostic process in this heterogeneous group, possibly containing also unrecognised NT2 or nonorganic hypersomnia, was not terminated and, therefore, no final diagnoses could be reached. This was substantiated by the frequent reclassification of diagnosis over time. Consequently, this mixed group of patients with different CDH would not have been suitable to reach our primary aim, namely to characterise typical SVT features within distinct CDH, and thus, was excluded from this study.

Insufficient sleep syndrome was diagnosed if patients with hypersomnolence (i) reported a shorter mean sleep duration than expected for their age, typically shorter on working days than on holiday; (ii) actigraphy showed shorter inactivity periods than expected, typically shorter on working than non-working days; or if (iii) prolongation of sleep duration improved their hypersomnolence. The term 'poor sleep hygiene' was used if either bedtimes or getting-up time greatly varied (more than two hours according to the medical history or actigraphy) in patients with hypersomnolence. If a patient was diagnosed with insufficient sleep syndrome and poor sleep hygiene, insufficient sleep syndrome was considered to be more important with respect to hypersomnolence and therefore defined as primary diagnosis.

The final clinical diagnosis was based on all information available, including SVT results analysed in this study, and is therefore subject to a certain circular reasoning.

## Extraction and eligibility criteria

The process of data extraction and selection and compilation of datasets is illustrated in figure 1. Encrypted data from 1997 to 2018 were extracted excluding recordings collected under treatment with stimulants (e.g. methylphenidate or modafinil) or continuous positive airway pressure, as well as assessments for fitness to drive. Recordings collected under treatment with non-stimulant drugs for mood disorders, neurodegenerative diseases (i.e. Parkinson's or dementia), restless legs syndrome, and epilepsy were marked but not excluded.

Data of patients with multiple consultations and/or paraclinical investigations were reviewed and the oldest recordings including an MSLT and a second SVT obtained within 3 months were selected, as long as the patient was not undergoing any kind of treatment.

In order to establish the most accurate and fact-based primary diagnosis, clinicians selected the diagnosis most likely explaining the patient's symptoms. The selection was based on history and any SVT results available in the clinical records.

Patients with multiple diagnoses were sorted in the following manner: the primary or main diagnosis was defined as the one most likely resulting in EDS; the secondary diagnosis was defined as the second most likely diagnosis for EDS; the tertiary diagnosis was defined as the most relevant among the remaining diagnoses. The likelihood of a diagnosis causing EDS was defined according to the clinical experience of the authors and the diagnoses of interest are listed with decreasing severity of

EDS in the following order: NT1, NT2, idiopathic hypersomnia, nonorganic hypersomnia, insufficient sleep syndrome, fatigue syndromes. Exceptions were only made if a more likely diagnosis was only suspected while a less likely diagnosis was confirmed. Sleep apnoea was only accepted as the primary diagnosis in case of an apnoea-hypopnoea index (AHI) ≥5/h in the absence of a diagnosis of interest. To reduce bias resulting from comorbidity, patients with multiple diagnoses of interest were included for the descriptive statistical analysis (dataset A) but excluded for further analysis (datasets B and C).

## Selection of diagnostic groups and SVT parameters of interest

NT1, NT2, idiopathic hypersomnia, insufficient sleep syndrome, and nonorganic hypersomnia are the most challenging sleep-wake disorders to diagnose, due to their ambiguity with similar signs and symptoms and were therefore selected as the diagnostic groups of key interest. Fatigue syndromes was selected in addition to CDH due to its important role as a differential diagnosis. Patients diagnosed with sleep apnoea as the primary cause for EDS served as comparison group to CDH.

For each SVT, only the most important parameters, considered to be potentially valuable in the diagnostic process, were analysed in this study, as listed in table 1. In order to reduce circular reasoning and potentially identify new differentiating variables, we deliberately excluded SOREMPs in the MSLT and PSG recordings.

## Statistical analysis

SPSS Statistics (IBM Corp. Released 2019. IBM SPSS Statistics for Macintosh, Version 26.0. Armonk, NY: IBM Corp.) was used for the statistical analysis.

To compare SVT results between diagnostic groups, a Kruskal-Wallis test was applied. Chi-Square ($\chi 2$) was reported as test statistic result, with $p < 0.05$ as the level for statistical significance. Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure and Bonferroni correction for multiple comparisons. The sample-size-adjusted effect size of each pairwise comparison was calculated using the standardised test statistic output (z-score). Results were reported as medians with interquartile range (IQR) as a measure of spread. Effect size was reported as Cohen's *r* (1988), describing effects as small (0.1 - 0.3), intermediate (0.3 - 0.5), and strong (≥ 0.5) [25]. For statistically significant pairwise comparisons between two diagnostic groups, receiver operating characteristics (ROC) curves were calculated. For each pairwise comparison, the three best discriminating SVT (highest effect size, largest area under the curve) and the corresponding optimal cut-off values were identified using the Index of Union (IU) method [26].

Dataset C was created by missing values imputation in dataset B, using the Estimated-Means method [27]. As exploratory analysis, all SVT parameters were correlated with each other (Pearson correlation matrix) and a 'TwoStep' cluster analysis was performed using SVT parameters and diagnostic categories as predictors. The silhouette measure of cohesion was used as measure of fit. A Multinomial Logistic Regression was performed to ascertain the weight of SVT parameters in the differentiation of diagnostic groups and the likelihood with which the diagnostic groups could be

predicted. Chi-Square ($\chi 2$) values of Likelihood Ratio Tests and Pseudo-$R^2$ (McFadden) measures were reported for overall goodness of fit. For all tests applied, statistical significance was defined as $p < 0.05$.

## Data availability

The ethical approval for this study enabled us to perform and publish a retrospective analysis of clinical patient data, however, we are not permitted to publish the clinical data set as such.

# Results

## Demographics (datasets A and B)

A demographic overview of the 1,352 patients is summarised in table 2. The most frequent primary diagnosis was sleep apnoea (33.8%), followed by insufficient sleep syndrome (15.2%), nonorganic hypersomnia (14.2%), fatigue syndromes (13.5%), idiopathic hypersomnia (8.8%), NT1 (7.8%), and NT2 (6.7%) (table 2).

Among patients primarily diagnosed with NT1, NT2, nonorganic hypersomnia, or insufficient sleep syndrome, the most common secondary diagnosis was sleep apnoea (14.1 - 22.9%). For idiopathic hypersomnia, the most frequent secondary diagnosis was nonorganic hypersomnia (14.3%) and fatigue syndromes patients were most often diagnosed with poor sleep hygiene as secondary diagnosis (23.5%). Please note that at the early stage of the diagnostic process the secondary diagnosis sometimes in fact represents a differential diagnosis to the primary diagnosis.

Pairwise comparisons (Kruskal-Wallis test) showed several significant differences in age and sex distribution between diagnostic groups in dataset B. Sleep apnoea patients were significantly older than all other groups except NT2, and idiopathic hypersomnia patients were significantly younger than all other groups. As expected from the official criteria for the final clinical diagnosis, two or more SOREMPs in the MSLT were most frequent in NT1 (86.5%) and in NT2 (84.3%) but rare (< 3%) in all other diagnostic groups. In sleep apnoea and insufficient sleep syndrome, the proportion of males was significantly higher and in idiopathic hypersomnia, nonorganic hypersomnia, and fatigue syndromes significantly lower compared to the proportion of females (table 2). Despite these significant differences across groups, age and sex did not substantially influence the overall diagnostic group prediction rate in the Multinomial Logistic Regression model (correct prediction of diagnostic group in 49.9% (+0.4%) of cases; see subchapter on the Multinomial Logistic Regression model, dataset C).

# Pairwise differences in SVT between patient groups (dataset B)

Excluding patients with multiple diagnoses of interest (*n* = 251) resulted in 1,101 patients and a total of 8,391 SVT, constituting dataset B (table 2b, table S1). SVT results for each patient group are presented in table 3 and figure 2. Even though a great overlap between patient groups exists in most SVT (figure 2), pairwise comparison between patient groups pointed to potentially valuable diagnostic differences (figure 3, table S2), except for NT1 and NT2 that did not significantly differ in any SVT. In the following subchapters, the results of the pairwise comparisons between the diagnoses of interests will be reported, excluding the results on the AHI and the comparisons between sleep apnoea and any other diagnosis which will be reported later in a separate subchapter. Mean & standard deviation are reported in table S3.

## MSLT, MWT, and ESS

A similar distribution of SLAT (in the MSLT) and the sleep latency in the MWT (WLAT) across the different CDH patient groups and fatigue syndromes was observed, with the ESS inversely following that pattern. For these SVT, the similar results in some patient groups enabled to subdivide CDH and fatigue syndromes into the following three 'categories': the narcolepsy category (NT1, NT2), the intermediate category (idiopathic hypersomnia, nonorganic hypersomnia, insufficient sleep syndrome), and fatigue syndromes as the third category. Within each of the three categories, no significant differences were found for SLAT, WLAT, or ESS.

The median SLAT and WLAT was short for the narcolepsy category, moderate for the intermediate category, and long for fatigue syndromes (table 3, figures 2&3). Similarly, but inverse, the median ESS scores were highest (≥ 16) for the narcolepsy category, moderate (12-13) for the intermediate category, and lowest (10) for fatigue syndromes (table 3, figures 2&3).

## SCT, PVT, and PSG (PSE)

No significant differences were found within the narcolepsy category. However, interesting deviances from the systematic differences in the MSLT, MWT, and ESS were found for the pairwise comparisons of the narcolepsy category with the intermediate category or fatigue syndromes, for comparisons within the intermediate category, and the comparison between the intermediate category and fatigue syndromes.

Interestingly, the median results of the narcolepsy category did not significantly differ from those of nonorganic hypersomnia in any of these three SVT parameters (table 3, figures 2&3). In addition, no significant difference was found between NT2 and idiopathic hypersomnia in the PVT in which the narcolepsy category showed the longest reaction time (PVTrt). In contrast, the median PSE differed between NT1 and idiopathic hypersomnia and also between NT2 and fatigue syndromes, but no significant difference was found for the other diagnoses. These differences may be explained by the fact that idiopathic hypersomnia was only diagnosed in the presence of a subjectively increased sleep need, and thus the median PSE was highest in idiopathic hypersomnia (followed by NT2) while it was lowest in fatigue syndromes and only slightly higher in NT1. The (non-significant) difference in

the median PSE between NT1 and NT2 may be explained by cases in the NT2 group resembling idiopathic hypersomnia. Narcolepsy with a begin in adolescence often starts as NT2, associated with excessive need for sleep, whereas later in the course of the disease the evolvement into NT1 goes along with maintenance insomnia [1, 28].

Within the intermediate category, nonorganic hypersomnia patients showed significantly higher error rates in the SCT (SCTer), longer PVTrt and a lower PSE compared to either idiopathic hypersomnia or insufficient sleep syndrome, or both (table 3, table S2, figures 2&3). This is in contrast to all other six SVT parameters, for which no significant differences were found within the intermediate category. Of note, nonorganic hypersomnia patients showed a disproportionately high SCTer compared to both idiopathic hypersomnia and insufficient sleep syndrome ($p < 0.001$), a longer PVTrt compared to insufficient sleep syndrome ($p < 0.001$), and a significantly lower PSE compared to idiopathic hypersomnia patients ($p < 0.001$) (table S2).

The pairwise comparison of SCT and PVT between fatigue syndromes and the intermediate category resulted in non-significant differences, except for the comparison with nonorganic hypersomnia where PVTrt was significantly longer ($p = 0.011$). The median PSE of fatigue syndromes patients was significantly lower compared to idiopathic hypersomnia and insufficient sleep syndrome while it did not differ from other diagnoses.

## Pupillography and actigraphy

Within the narcolepsy category and the intermediate category, no significant differences were found for the median PUI and actigraphy index (ACTI). Interestingly, pairwise comparisons for PUI and ACTI remained mostly non-significant across all diagnoses of interest. The median PUI was highest in the narcolepsy category with a significant difference compared only to the nonorganic causes nonorganic hypersomnia and fatigue syndromes. The median ACTI was highest in fatigue syndromes and significantly lower in NT1 and insufficient sleep syndrome patients.

## AHI and pairwise comparison of sleep apnoea with CDH and fatigue syndromes

The highest median AHI was found in sleep apnoea (20/h), which is not surprising since an AHI < 5/h could, by definition, not occur in sleep apnoea patients. The median AHI of all other diagnoses was significantly lower (< 5/h) and no significant difference was found between those diagnoses (table 3, figure 3).

The pairwise comparison with the narcolepsy category resulted in significantly longer median SLAT and WLAT, lower ESS scores, and lower PUI for sleep apnoea. The median SCTer and PSE in sleep apnoea were lower than the ones in the narcolepsy category, but the difference was only significant for NT1 (SCTer) and NT2 (PSE) respectively.

For the comparison of sleep apnoea with the intermediate category, the median differences of SLAT and WLAT were non-significant while the median PSE was significantly lower in sleep apnoea. SCTer was higher in sleep apnoea compared to idiopathic hypersomnia and insufficient sleep syndrome but

not nonorganic hypersomnia. Furthermore, median ESS scores, PUI, and ACTI were significantly lower in sleep apnoea compared to idiopathic hypersomnia and PVTrt was significantly higher in sleep apnoea compared to insufficient sleep syndrome.

Pairwise comparisons between sleep apnoea and fatigue syndromes revealed non-significant results for ESS, PSE, and PUI while SLAT and WLAT were shorter, SCTer higher, PVTrt longer, and ACTI lower in sleep apnoea.

## Best discriminating SVT and cut-off values (dataset B)

In a first step, among all pairwise comparisons across all SVT, only those resulting in significant differences were identified. In a second step, they were ranked for each pairwise comparison according to their effect size. As a result, the best three discriminating SVT parameters for each pairwise comparison are reported in table 4. In addition, the corresponding cut-off values resulting in the best combination of sensitivity and specificity are reported. Excluding sleep apnoea, the best discriminating SVT parameters among the 15 pairs of diagnoses of CDH and fatigue syndromes were SLAT (7 pairs), WLAT (4 pairs), and PSE / PVTrt (each 1 pair). For discrimination between NT1 and NT2 and between idiopathic hypersomnia and insufficient sleep syndrome, no SVT was found. For second- and third-ranking SVT see table 4.

After exclusion of sleep apnoea and AHI, among the best three discriminating SVT parameters with decreasing frequency were WLAT (11 pairs) and SLAT (10 pairs), ESS (6 pairs), SCTer (5 pairs), PSE (2 pairs), and PVTrt, ACTI and PUI (1 pair each).

The range or single value of the optimal cut-off values for those parameters, as listed in table 4, were: WLAT = 17.6 – 35.4 min, SLAT = 3.6 – 9.4 min, ESS = 11.5 – 15.5, SCTer = 3 – 3.5%, PSE = 0.93 – 0.94, PVTrt = 284 ms, ACTI = 35 %, and PUI = 8.8. A broad range suggests that different cut-off values are required for many pairs of patient groups.

The highest sensitivity and specificity (both ≥ 0.95) was found for the discrimination of NT1 and fatigue syndromes by SLAT, while the lowest sensitivity and specificity (both < 0.65) was found for the discrimination of idiopathic hypersomnia and nonorganic hypersomnia by PSE or SCTer.

## Cluster analysis (dataset C)

A total of 1,518 (15.3%) missing values were imputed to form a complete dataset with 9,909 SVT results. The mean, standard deviation, and distribution of SVT results among patient groups remained statistically identical (table S4). Correlations between SVT parameters are illustrated in a correlation matrix (table S5). WLAT was the SVT for which the most significant correlation coefficients > |0.3| were found in comparison to other SVT parameters.

The initial 'TwoStep' cluster analysis, with automatic selection of best fitting cluster count, resulted in two clusters: those with and those without sleep apnoea. Therefore, we decided to run a second

'TwoStep' cluster analysis excluding patients with sleep apnoea. This analysis resulted in a fair model quality (silhouette measure of cohesion = 0.2). Four clusters were formed as 'best model fit'. They used the parameters in the following decreasing order of importance: SCTer, PVTrt, WLAT, SLAT, PSE, diagnostic group, ESS, ACTI, and PUI. The four clusters consisted of the following diagnostic group proportions: Cluster one = FS (52%) + ISS (20%) + NOH (17%) + IH (9%) + NT1 (1%) + NT2 (1%); cluster two = ISS (28%) + NOH (23%) + IH (17%) + NT1 (14%) + NT2 (12%) + FS (6%); cluster three = FS (49%) + NOH (24%) + IH (13%) + ISS (12%) + NT1 (1%) + NT2 (1%); cluster four = NT1 (38%) + NT2 (29%) + NOH (16%) + FS (8%) + ISS (6%) + IH (3%).

## Prediction within multiple diagnostic groups (dataset C)

After the missing value imputation, the Multinomial Logistic Regression model showed a good and significant overall goodness of fit, with a Likelihood Ratio Test $\chi2(54) = 1761.1$, $p < 0.001$, and Pseudo-$R^2$ (McFadden) = 0.458. All SVT parameters were statistically significant ($p < 0.05$) while AHI ($\chi2(6) = 1049.1$, $p < 0.001$) and SLAT ($\chi2(6) = 148.4$, $p < 0.001$) were by far the strongest weighted parameters. Other SVT parameters were weighted by the model in the following decreasing order: PVTrt, WLAT, ACTI, ESS, PSE, PUI, SCTer. The best prediction of the model resulted for sleep apnoea (94.7%), while it predicted the correct diagnosis in 64.6% of cases overall (table 5). The subsequent exclusion of sleep apnoea and AHI did not reveal a relevant impact on the Multinomial Logistic Regression model (Likelihood Ratio Test $\chi2(40) = 623.9$, $p < 0.001$, Pseudo-$R^2$ (McFadden) = 0.261). Again, all SVT variables were statistically significant ($p < 0.05$) and SLAT resulted as the most important parameter ($\chi2(5) = 146.0$, $p < 0.001$), other SVT parameters were weighted in the same order as mentioned above. Overall, the model predicted the correct diagnosis in 49.5% of cases (table 5). The best prediction was obtained for fatigue syndromes (78.0%) and NT1 (62.2%). In both models, idiopathic hypersomnia was least likely to be correctly predicted (19.3 %) followed by NT2 (23.9%).

# Discussion

This is the first study which systematically analysed the diagnostic value of a great number of SVT in a large clinical population over a long period of time. A total of 8,391 SVT data from 1,101 patients diagnosed with CDH (NT1, NT2, idiopathic hypersomnia, nonorganic hypersomnia, insufficient sleep syndrome), fatigue syndromes, or sleep apnoea were retrospectively analysed (table 3). Since by definition, sleep apnoea could be clearly differentiated from the other diagnoses by the AHI (figure 2) this patient group was treated separately.

The first aim of the study was to characterise the spectrum between sleepiness, hypersomnia, tiredness, and fatigue by analysing a variety of SVT assessing sleep propensity, sleep duration, sleep quality, and the capacity to counteract sleepiness.

In the case of multiple diagnoses in a patient, the primary diagnosis was selected based on a ranking of diagnoses (and categories) in terms of EDS severity, which was defined according to clinical experience in the following descending order: NT1, NT2 (narcolepsy category), idiopathic

hypersomnia, nonorganic hypersomnia, insufficient sleep syndrome (intermediate category), and fatigue syndromes (third category). The explorative analysis showed the same order for WLAT (short to long), ESS (high to low), and PUI (high to low). Except for the median SLAT in insufficient sleep syndrome, which was non-significantly shorter than the one of idiopathic hypersomnia and nonorganic hypersomnia, the same order also applied for SLAT (short to long).

Rather unexpectedly, a short SLAT was present in the majority of patients with nonorganic hypersomnia (table 3, table S6). This underlines the fact, that at least a subgroup of patients with a psychiatric disorder and a subjective complaint of EDS or prolonged sleep need is indeed objectively sleepy in the course of their illness.

The narcolepsy category differed from the intermediate category and fatigue syndromes, and the intermediate category also differed from fatigue syndromes (figure 3) in most SVT. No differences were found in this respect between NT1 and NT2. Within the intermediate category, significant differences were found for pairwise comparisons with nonorganic hypersomnia, however, only for SCTer, PVTrt, and PSE.

Based on pairwise comparisons across diagnoses (excluding sleep apnoea), the most important SVT parameters were identified, and the corresponding cut-off values calculated (table 4). No single SVT was able to differentiate between all patient groups, which is explained by the considerable overlap of SVT results among all patient groups (figure 2). However, specific SVT were able to differentiate between specific pairs of patient groups with great sensitivity and specificity. Among the top three SVT variables for each pairwise comparison, WLAT and SLAT proved to be most important, followed by ESS, SCTer, PSE, PVTrt, ACTI, and PUI.

In the cluster analysis after exclusion of patients with sleep apnoea, the four clusters formed were somewhat comparable to the three diagnostic categories (narcolepsies in cluster four, intermediate category in cluster two and fatigue syndromes in cluster one and three), but much less to the established diagnostic groups. Therefore, the identified clusters do not help to differentiate the most ambiguous diagnoses. Interestingly, the most important parameters forming the clusters were the SCTer and PVTrt, followed by WLAT and SLAT.

The Multinomial Logistic Regression model predicted the correct diagnosis in only 49.5% of patients (excluding patients with sleep apnoea and the AHI parameter). The mathematical model had a particularly poor prediction rate for idiopathic hypersomnia (19.3%), NT2 (23.9%), and nonorganic hypersomnia with prolonged sleep need (31.2 %).

A well-established SVT variable for the differentiation of CDH (and fatigue syndromes) is the number or the latency of SOREMPs [29]. However, the comparison of SVT results with a ‘clinical’ diagnosis which also partly relies on SVT results obviously leads to a certain degree of circular reasoning, which we aimed to avoid as far as possible. Therefore, we excluded SOREMs as an SVT variable and used it only for the clinical diagnosis.

While the present SVT results did not allow discrimination between NT1 and NT2, rather strong diagnostic criteria for NT1 are often found (cataplexy, low hypocretin) which have not been taken into account in this study [1]. In contrast, the diagnostic criteria of NT2 and even its existence and differentiation from idiopathic hypersomnia are subject to an ongoing and controversial debate [7,

30]. The most ambiguous diagnoses among CDH, and therefore most difficult to identify, are probably NT2, idiopathic hypersomnia, nonorganic hypersomnia, and insufficient sleep syndrome.

The most prominent SVT parameter discriminating NT2 from idiopathic hypersomnia, nonorganic hypersomnia, and insufficient sleep syndrome was a WLAT < 19 min (figure 2, table 4), a tool that has not been broadly used in clinical work-up for discriminating CDH so far. Idiopathic hypersomnia and insufficient sleep syndrome are the most difficult diagnoses to differentiate by SVT results alone, suggesting a similar pathogenesis. Idiopathic hypersomnia could theoretically represent a more pronounced form of insufficient sleep syndrome in patients with an even greater need for sleep. The rather well preserved capacity to remain awake in the MWT, found in idiopathic hypersomnia and insufficient sleep syndrome compared to NT1 and NT2, could suggest a different pathology. We can only speculate, that the narcolepsy category is affected by both an increased sleepiness level but also by an impaired capacity to maintain wakefulness, while the latter is preserved in insufficient sleep syndrome and idiopathic hypersomnia. The PSE was highest in patients with idiopathic hypersomnia but did not significantly differ from the ones with insufficient sleep syndrome, nor did any other SVT between those two diagnoses. In clinical practice, both diagnoses are usually differentiated by instructing the patient to extend the sleep period, since sleepiness in insufficient sleep syndrome resolves when extending the total sleep time while it does not in idiopathic hypersomnia [1, 31].

Patients with nonorganic hypersomnia, a term used as a synonym of hypersomnia due to psychiatric disorder (depression, anxiety, attention deficit hyperactivity disorder etc.), are unfortunately only rarely included in larger studies. Consequently, nonorganic hypersomnia is poorly investigated by objective measurements and often not included in reviews of CDH [32, 33]. Nevertheless, nonorganic hypersomnia represents a common differential diagnosis within CDH [34]. In this study, several SVT parameters were found useful for discriminating nonorganic hypersomnia among other ambiguous diagnoses: The SCTer was disproportionately high in nonorganic hypersomnia compared to idiopathic hypersomnia and insufficient sleep syndrome, the PVTrt longer compared to insufficient sleep syndrome, and the PSE lower compared to idiopathic hypersomnia (figure 2, table 4). We speculated earlier that the disproportionately poor performance in the active performance tests of patients with nonorganic hypersomnia (compared to SLAT and WLAT) is not related to true sleepiness but rather to a reduced amount of psychic energy and motivation in patients with a nonorganic disorder [35]. The better discriminating power of the SCT compared to the PVT is probably related to its ‘go/no go’ principle, requiring a greater vigilance level compared to the simple reaction task in the PVT. Due to clinical similarities with CDH, we also included fatigue syndromes which could be differentiated primarily by a long SLAT (> 9.4 min; figure 2, table 4).

It could be argued, that within the spectrum of CDH the differentiation between diagnostic groups has not the same impact on treatment as for e.g. sleep apnoea or insufficient sleep syndrome. This might be true for drug treatment, except for the differentiation of nonorganic hypersomnia from other disorders. However, the correct differentiation between e.g. NT2 and idiopathic hypersomnia might influence behavioural treatment (napping method) and may affect the difficulties for reimbursement by the insurances.

In addition to the pairwise comparison of SVT, the present study analysed if a combination of SVT could improve the discrimination between diagnoses. From our clinical experience, we have

postulated many years ago that the aim should not be to find 'the one' best SVT to assess sleepiness, but rather to find a combination of SVT tailored to the specific clinical problem [36]. This proposition was based on the assumption that sleepiness is a multi-dimensional construct and CDH can differ in various dimensions, between sleepiness (~ sleep propensity), tiredness (loss of energy), and fatigue (time on task performance decrement) [37].

The results of this study suggest that a 'tailored' selection of important SVT variables could support clinicians in the differential diagnostic process. However, it might be difficult to create an overall model for the identification of the correct diagnosis based on such a selection, especially for the most ambiguous diagnoses. The additional inclusion of a greater number of clinical and paraclinical data ('big data') could most certainly deliver better results and support particularly clinicians with less experience in sleep-wake medicine [38]. It is open to debate if such an approach could outperform the diagnostic skills of an experienced sleep-wake clinician, who recognises patterns of many subtle symptoms and signs in the diagnostic process leading to the final diagnosis.

The present study made a first step towards a hybrid approach, which relies on the clinician to, firstly, narrow down the differential diagnosis, secondly, to consult the corresponding suggested SVT results, and finally, to verify the suggested diagnosis in the clinical context.

We were able to confirm our hypothesis that a combination of multiple SVT assessing not only sleep propensity but also the capacity to counteract EDS was superior compared to the use of a single SVT. The best discrimination power of the SVT battery was achieved when combined with clinical judgement in order to reduce the differentiation to the pair of most probable CDH diagnoses (table 4).

One of the main strengths of this study is the data set. In the Sleep-Wake-Epilepsy-Centre in Bern, we aimed at a multimodal diagnostic vigilance battery early on and started in 2002 to systematically add the MWT and various performance tests to the standard assessment of patients with suspected CDH (PSG, MSLT, and actigraphy) for the following reason: We and others recognised that the capacity to remain awake despite sleepiness or tiredness as measured e.g. by the MWT does not merely reflect the reciprocal of sleep propensity as measured by the MSLT [36, 39]. Therefore, a great number of SVT were not only used for treatment control or to assess fitness to drive but also as diagnostic tools. However, such an extensive vigilance battery has not been widely used so far in the clinical diagnostic process, mainly because of the large resources such a battery requires.

## Limitations

The present analysis provides the approximate diagnostic values of a limited number of SVT parameters in the work-up of patients suspected with CDH, diagnosed according to the various ICSD versions. Since only a highly specific selection of relatively few parameters of available SVT parameters were analysed, the inclusion of a wider spectrum of SVT and more (para-)clinical parameters may increase the diagnostic power even more. Also, the present calculations of SVT parameter cut-off values are just a first attempt and have not yet been verified in other datasets. The selection of SVT was based on long-standing clinical expertise, particularly when assessing fitness to drive. However, the number of tests that could be included in the clinical routine protocol

was limited. For example the SCT and the sustained attention to response task (SART) [40], represent a similar go/no-go paradigm, however, only one of them could be selected.

The diagnostic value of the MSLT and MWT beyond sleep latency and SOREMPs could be increased by analysing the individual naps and adding also the sleep efficiencies within each nap to get more information about sleep propensity during the day. Moreover, it has been shown previously by several authors that the sequence of sleep stages before rapid eye movement onset and the precise way of falling asleep contains important information of diagnostic power [41, 42].

The limited sleeping time during clinical routine polysomnography and use of the average inactivity index over a two-week period of actigraphy do not allow thorough assessments of sleep need and rest duration. Measurements of sleep duration and fragmentation beside efficiency, would be valuable during 'ad libitum' polysomnography but due to time and cost constraints, it is impossible to systematically perform this type of polysomnography in the clinical setting. The comparison of the inactivity index between workdays and weekends or holidays is much more feasible and could eventually enhance the diagnostic value of the actigraphy [43, 44]. However, the differentiation of sleep and awake inactivity outside of polysomnography remains a challenge and clinical studies with mobile applications or wearables are pending [45]. Therefore, we defined prolonged sleep time using the number of night-sleep hours during days off (e.g. holidays) indicated in the medical history, with a cut-off value of > 10 hours.

We must acknowledge that the gold standard chosen here, "final clinical diagnosis", is rarely truly final. And the diagnostic classifications will stand only until they are replaced due to new insights. As already stated by Bedrich Roth in 1962, "we must constantly remember that any classification is basically an artificial simplification of reality, which should not become a brake for scientific progress" [8]. Such reflections have most recently culminated in a promising proposition for a new diagnostic schematic of CDH [6, 7]. The results of the initial MSLT included in this study did not always fit the later final diagnosis which was occasionally only confirmed by a subsequent MSLT. This explains why SLAT in NT2 and idiopathic hypersomnia was sometimes longer than the diagnostic limit of eight minutes (table S6). Furthermore, it is in line with evidence that when testing repetitively, SLAT can vary, which is of diagnostic relevance in particular for NT2 and idiopathic hypersomnia [46].

The rather limited impact on correct classification by the SVTs and the outcome of only four clusters (after exclusion of sleep apnoea) in the cluster analysis could support the view of many sleep experts, that the current classification based on clearly defined entities might be problematic. The great overlap of SVT parameters would probably better fit to a concept of a 'hypersomnolence spectrum disorder' with individually varying weights of the multiple underlying comorbidities, including e.g. genetic, psychological and life-style factors. We agree with Lammers et. al, that the multidimensional aspects of hypersomnolence, complemented by levels of certainty, would profit from a 'pattern recognition based diagnostic process' which would optimise clinical care, tailored to individual patients [6]. An example of this great overlap of symptoms can be observed in narcolepsy patients with onset during adolescence, in which many aspects of idiopathic hypersomnia can be found such as prolonged sleep duration and non-restorative naps [28].

The important overlap of SVT results between idiopathic and nonorganic hypersomnia indicates that patients with nonorganic hypersomnia can be objectively sleepy. It can be speculated that these

similarities could be explained by a common risk factor, i.e. 'long-sleeper type', often present already in childhood. Future long-term studies should be performed, to clarify if patients with psychiatric disorders who have described themselves as 'short sleepers' preferentially develop insomnia while 'long sleepers' more often develop hypersomnolence.

It should be underlined that the diagnosis 'nonorganic hypersomnia' does not implicate absence of any organic cause in a narrower sense. In patients with nonorganic hypersomnia, hypersomnolence may be particularly resistant to antidepressive treatment [35, 47], and therefore, the term 'nonorganic hypersomnia' simply stands for a temporary association of hypersomnolence with a psychiatric disorder or psychiatric symptoms. Consequently, an increased depression score and/or a current psychiatric diagnosis (e.g. depression) was not required for diagnosing nonorganic hypersomnia. This is one reason explaining why depression as a secondary or tertiary diagnosis occurred rather rarely in our population of patients with nonorganic hypersomnia. Another reason is the referral bias. Patients with a clear diagnosis of depression and EDS are generally not referred to a sleep laboratory because the aetiology of their sleepiness is not questioned and the MSLT is not mandatory for diagnosis (nor would it be reimbursed by the health insurance in Switzerland).

Furthermore, the combination of multiple co-morbidities in the same patient inevitably will affect diagnostic values of SVT and other parameters, ultimately requiring a more complex mathematical model. An organic handicap such as idiopathic hypersomnia or narcolepsy often causes later evolving psychiatric disorders such as depression, resulting in a mixture of symptoms and signs as exemplified in a survey with subjective scales by Drooglever et al. [48]. In such patients, objective biological findings would be particularly helpful to allow the diagnosis of both underlying disorders.

A limitation of this study is, that idiopathic hypersomnia was restricted to those with prolonged sleep need because, for more than two decades in our centre, patients not complaining of prolonged sleep need were diagnosed with 'EDS of unknown origin'. Among others, this included patients with idiopathic hypersomnia without prolonged sleep need, characterised by EDS but not hypersomnia in a strict sense. This mixed group of 'EDS of unknown origin' was not suitable for this study with the primary aim of investigating disease-specific SVT characteristics. However, this mixed group is of very high interest for future preferentially prospective studies that focus on the identification of new disease clusters using an optimised and extended selection of clinical and paraclinical biomarkers (incl. SVT) [49].

Finally, a certain degree of circular reasoning has to be taken into account, since the final clinical diagnoses recorded in the Bern Sleep Database were based on any available information on a given patient to the unblinded clinician, including the SVT results. The presence of cut-off values derived from SVT, e.g. AHI or SLAT, certainly had a major impact on some diagnoses such as sleep apnoea, NT1, NT2, or idiopathic hypersomnia.

# Conclusion

The optimal SVT parameters should be carefully selected in order to allow a reliable differentiation between the two most probable diagnoses as suspected from clinical judgement. An individual test or a combination of a limited number of SVT parameters cannot differentiate reliably between all diagnostic groups. Our findings underline the importance of the MSLT as most valuable SVT in the diagnostic work-up battery of CDH but suggests adding further SVT. In particular, the MWT could further improve the diagnostic accuracy, especially when differentiating between the narcolepsies and other CDHs with severe daytime sleepiness (idiopathic hypersomnia, nonorganic hypersomnia, and insufficient sleep syndrome). The MWT results could be diagnostically helpful in patients with NT2 or IH showing only one SOREMP in the MSLT. Additionally, the inclusion of a performance-based vigilance test such as the SCT may be helpful for the differentiation of nonorganic hypersomnia from idiopathic hypersomnia or insufficient sleep syndrome.

# Acknowledgements

We are grateful to the numerous residents and senior physicians who carefully added these data systematically into the Bern Sleep Database from 1997 to 2018 and the technical and secretary staff of the Sleep-Wake-Epilepsy-Centre in the Inselspital for their longstanding support. We thank Andrea Maria Mathis for her help in data transformation.


# Disclosure statement

Financial Disclosure: none

Non-financial Disclosure: none

Preprint repositories: none

# References

1. The International Classification of Sleep Disorders. Third Edition. Darien, IL, American Academy of Sleep Medicine; 2014.

2. Kretzschmar U, Werth E, Sturzenegger C, Khatami R, Bassetti CL, Baumann CR. Which diagnostic findings in disorders with excessive daytime sleepiness are really helpful? A retrospective study. J Sleep Res. 2016; 25 (3): 307-313. doi: 10.1111/jsr.12383.

3. Ohayon MM, Dauvilliers Y, Reynolds CF. Operational definitions and algorithms for excessive sleepiness in the general population: Implications for DSM-5 nosology. Arch Gen Psychiatry. 2012; 69 (1): 71-79. doi: 10.1001/archgenpsychiatry.2011.1240.

4. Ohayon MM. From wakefulness to excessive sleepiness: What we know and still need to know. Sleep Med Rev. 2008; 12 (2): 129-141. doi: 10.1016/j.smrv.2008.01.001.

5. Sommerauer M, Valko PO, Werth E, Baumann CR. Excessive sleep need following traumatic brain injury: A case-control study of 36 patients. J Sleep Res. 2013; 22 (6): 634-639. doi: 10.1111/jsr.12068.

6. Lammers GJ, Bassetti CL, Dolenc-Groselj L, et al. Diagnosis of central disorders of hypersomnolence: A reappraisal by European experts. Sleep Med Rev. 2020; 52: 101-306. doi: 10.1016/j.smrv.2020.101306.

7. Fronczek R, Arnulf I, Baumann CR, Maski K, Pizza F, Trotti LM. To split or to lump? Classifying the central disorders of hypersomnolence. Sleep. 2020; 43 (8): 1-8. doi: 10.1093/sleep/zsaa044.

8. Roth B. ed. Narkolepsie und Hypersomnie, vom Standpunkt der Physiologie des Schlafes. Berlin, VEB Verlag Volk und Gesundheit; 1962.

9. Avlund K, Rantanen T, Schroll M. Tiredness and subsequent disability in older adults: The role of walking limitations. J Gerontol A Biol Sci Med Sci. 2006; 61 (11): 1201-1205. doi: 10.1093/gerona/61.11.1201.

10. Brown ID. Driver fatigue. Hum Factors. 1994; 36 (2): 298-314. doi: 10.1177/001872089403600210.

11. Phillips RO. A review of definitions of fatigue - And a step towards a whole definition. Transportation Research Part F: Traffic Psychology and Behaviour. 2015; 29: 48-56. doi: 10.1128/JCM.02231-09.

12. Shen J, Barbera J, Shapiro CM. Distinguishing sleepiness and fatigue: Focus on definition and measurement. Sleep Med Rev. 2006; 10 (1): 63-76. doi: 10.1016/j.smrv.2005.05.004.

13. Bigland-Ritchie B, Rice CL, Garland SJ, Walsh ML. Task-dependent factors in fatigue of human voluntary contractions. Adv Exp Med Biol. 1995; 384: 361-80. doi: 10.1007/978-1-4899-1016-5_29.

14. Oonk M, Davis CJ, Krueger JM, Wisor JP, Van Dongen HPA. Sleep deprivation and time-on-task performance decrement in the rat psychomotor vigilance task. Sleep. 2015; 38 (3): 445-451. doi: 10.5665/sleep.4506.

15. Hulst M Van Der, Meijman T, Rothengatter T. Maintaining Task set under Fatigue: a Study of time on Task effects in simulated Driving. Transportation Research Part F: Traffic Psychology and Behaviour. 2001; 4: 103-118. doi.org/10.1016/S1369-8478(01)00017-1.

16. Mathis J, Hatzinger M. Praktische Diagnostik bei Müdigkeit / Schläfrigkeit. Schweizer Archiv für Neurologie und Psychiatrie. 2011; 162 (8): 300-309.

17. Johns MW. A new method for measuring daytime sleepiness: The Epworth sleepiness scale. Sleep. 1991; 14 (6): 540-545. doi: 10.1093/sleep/14.6.540.

18. Wise MS. Objective measures of sleepiness and wakefulness: Application to the real world? J Clin Neurophysiol. 2006; 23 (1): 39-49. doi: 0.1097/01.wnp.0000190416.62482.42.

19. Wilhelm B, Wilhelm H, Lüdtke H, Streicher P, Adler M. Pupillographic assessment of sleepiness in sleep-deprived healthy subjects. Sleep. 1998; 21 (3): 258-265.

20. Mathis J, de Lacy S, Roth C., Hill EA. Measuring – Monitoring Sleep and Wakefulness. In: Bassetti C, McNicholas W, Paunio T, Peigneux P, eds. Sleep Medicine Textbook, 2nd Ed. European sleep research society; 2021: Chapter B3,181–200.

21. Oswald-Fink BF. Die Bedeutung der Vigilanztests in der multimodalen Diagnostik exzessiver Tagesschläfrigkeit. Diploma thesis. The University of Bern, Department of Psychology. 2007.

22. Findley LJ, Fabrizio MJ, Knight H, Norcross BB, LaForte AJ, Suratt PM. Driving simulator performance in patients with sleep apnea. Am Rev Respir Dis. 1989;140 (2): 529-30. doi: 10.1164/ajrccm/140.2.529.

23. American Academy of Sleep Medicine. The International Classification of Sleep Disorders. second Edition. Westchester, IL, American Academy of Sleep Medicine; 2005.

24. Association of Sleep Disorders Centers and the Association for the Psychophysiological Study of Sleep. Diagnostic Classification of Sleep and Arousal Disorders, first Ed. Sleep, 1979; 2: 5-137. doi.org/10.1093/sleep/2.1.1.

25. Cohen J. Statistical Power Analysis for the Behavioral Sciences. Hillsdale, NJ: Lawrence Erlbaum Associates; 1988.

26. Unal I. Defining an optimal cut-point value in ROC analysis: An alternative approach. Comput Math Methods Med. 2017. doi: 10.1155/2017/3762651.

27. University of Sussex online archive. IBM SPSS Missing Values 22 Handbook. http://www.sussex.ac.uk/its/pdfs/SPSS_Missing_Values_22.pdf. Accessed April 3, 2021.

28. Nevšímalová S. Disorders Associated with Increased Sleepiness. In: Nevšimalová S, Bruni B, eds. Sleep Disorders in Children. Springer International, Switzerland, 2017: Chapter 13, 281-304.

29. Zhang Z, Mayer G, Dauvilliers Y, et al. Exploring the clinical features of narcolepsy type 1 versus narcolepsy type 2 from European Narcolepsy Network database with machine learning. Sci Rep. 2018; 8(1). doi: 10.1038/s41598-018-28840-w.

30. Bassetti CL, Adamantidis A, Burdakov D, et al. Narcolepsy — clinical spectrum, aetiopathophysiology, diagnosis and treatment. Nat Rev Neurol. 2019; 15 (9): 519-539. doi: 10.1038/s41582-019-0226-9.

31. Baumann-Vogel H, Hoff S, Valko PO, Poryazova R, Werth E, Baumann CR. Extending sleep to confirm insufficient sleep syndrome is challenging. J Sleep Res. 2020. doi: 10.1111/jsr.13109.

32. Barateau L, Lopez R, Franchi JAM, Dauvilliers Y. Hypersomnolence, Hypersomnia, and Mood Disorders. Curr Psychiatry Rep. 2017; 19 (2). doi: 10.1007/s11920-017-0763-0.

33. Baumann CR, Mignot E, Lammers GJ, et al. Challenges in diagnosing narcolepsy without cataplexy: A consensus statement. Sleep. 2014; 37 (6): 1035-1042. doi: 10.5665/sleep.3756.

34. Plante DT. Sleep propensity in psychiatric hypersomnolence: A systematic review and meta-analysis of multiple sleep latency test findings. Sleep Med Rev. 2017; 31: 48-57. doi: 10.1016/j.smrv.2016.01.004.

35. Kofmel NC, Schmitt WJ, Hess CW, Gugger M, Mathis J. Sleepiness and performance is disproportionate in patients with non-organic hypersomnia in comparison to patients with narcolepsy and mild to moderate obstructive sleep apnoea. Neuropsychobiology. 2014; 70 (3): 189-194. doi: 10.1159/000365486.

36. Mathis J, Hess CW. Vigilance tests in narcolepsy. In: Bassetti CL, Billiard M, Mignot E, eds. Narcolepsy and Hypersomnia. New York, London, Informa Healthcare; 2007: 243-256.

37. Cluydts R, De Valck E, Verstraeten E, Theys P. Daytime sleepiness and its evaluation. Sleep Med Rev. 2002; 6 (2): 83-96. doi: 10.1053/smrv.2002.0191.

38. Šonka K, Šusta M, Billiard M. Narcolepsy with and without cataplexy, idiopathic hypersomnia with and without long sleep time: A cluster analysis. Sleep Med. 2015; 16 (2): 225-231. doi: 10.1016/j.sleep.2014.09.016.

39. Sangal RB, Thomas L, Mitler MM. Disorders of excessive sleepiness; Treatment improves ability to stay awake but does not reduce sleepiness. Chest. 1992; 102 (3): 699-703. doi: 10.1378/chest.102.3.699.

40. Robertson IH., Manly T, Andrade J, Baddeley BT, & Yiend, J. “Oops!”: Performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. Neuropsychologia, 35, 747–758. doi: 10.1016/s0028-3932(97)00015-8.

41. Liu Y, Croft JB, Wheaton AG, et.al. Association between perceived insufficient sleep, frequent mental distress, obesity and chronic diseases among US adults, 2009 behavioral risk factor surveillance system. BMC Public Health. 2013; 13: 84. doi: 10.1186/1471-2458-13-84.

42. Pizza F, Vandi S, Detto S, et al. Different sleep onset criteria at the multiple sleep latency test (MSLT): An additional marker to differentiate central nervous system (CNS) hypersomnias. J Sleep Res. 2011; 20: 250-256. doi: 10.1111/j.1365-2869.2009.00808.x.

43. Smith MT, McCrae CS, Cheung J, 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. doi: 10.5664/jcsm.7228.

44. Evangelista E, Lopez R, Barateau L, et al. Alternative diagnostic criteria for idiopathic hypersomnia: A 32-hour protocol. Ann Neurol. 2018; 83 (2): 235-247. doi: 10.1002/ana.25141.

45. Penzel T, Schöbel C. Other tests and examinations. In: Bassetti C, McNicholas W, Paunio T, Peigneux P, eds. Sleep Medicine Textbook, 2nd Ed. European sleep research society; 2021: Chapter B4, 201-207.

46. Lopez R, Doukkali A, Dauvilliers Y, et al. Test-Retest Reliability of the Multiple Sleep Latency Test in Central Disorders of Hypersomnolence. Sleep 2017; 40 (12). DOI: 10.1093/sleep/zsx164.

47. Baldwin DS, Papakostas GI. Symptoms of fatigue and sleepiness in major depressive disorder. J Clin Psychiatry. 2006; 67 Suppl 6: 9–15. PMID: 16848671.

48. Droogleever Fortuyn HA, Fronczek R, Smitshoek M, et al. Severe fatigue in narcolepsy with cataplexy. J Sleep Res. 2012; 21 (2): 163-169. doi: 10.1111/j.1365-2869.2011.00943.x

49. Dietmann A, Wenz E, van der Meer J, et al. The Swiss Primary Hypersomnolence and Narcolepsy Cohort study (SPHYNCS): Study protocol for a prospective, multicentre cohort observational study. J Sleep Res. 2021 30(5):e13296. doi: 10.1111/jsr.13296.

# Figure captions

**Figure 1 Flowchart illustrating data extraction, selection, and analysis.** Central disorders of hypersomnolence (CDH): narcolepsy with (NT1) and without cataplexy (NT2), idiopathic hypersomnia (IH), nonorganic hypersomnia (NOH), insufficient sleep syndrome (ISS) and fatigue syndromes (FS). Abbreviations: Multiple Sleep Latency Test (MSLT); Sleep and Vigilance Test (SVT); sleep apnoea (SA).

**Figure 2 Boxplots of Sleep and Vigilance Test results.** Narcolepsy with (NT1) and without cataplexy (NT2), idiopathic hypersomnia (IH), nonorganic hypersomnia (NOH), insufficient sleep syndrome (ISS), fatigue syndromes (FS), and sleep apnoea (SA). Boxplots consist of interquartile range (IQR: 25 to 75 % or Q1 to Q3; box), median (line in box), whiskers (minimum: Q1 - 1.5*IQR, maximum: Q3 + 1.5*IQR), and outliers below minimum or above maximum (circles = outliers (>1.5*IQR), stars = extreme values (>3.0*IQR)).

**Figure 3 Effect size of pairwise comparisons in Sleep and Vigilance Tests.** The sample-size-adjusted effect size of each pairwise comparison was calculated using the standardised test statistic output (z-score), in order to determine which pairwise differences were most relevant within sleep and vigilance tests (SVT) and in order to compare differences between diagnostic groups among SVT. Effect size (rounded, for unrounded values see supplementary table S2) was reported as Cohen's r (1988), where the following intervals are reported: 0.1 to 0.3: small effect; 0.3 to 0.5: intermediate effect; 0.5 and higher: strong effect. *: Count of significant pairwise comparisons. Abbreviations: Narcolepsy with (NT1) and without cataplexy (NT2), idiopathic hypersomnia (IH), nonorganic hypersomnia (NOH), insufficient sleep syndrome (ISS), fatigue syndromes (FS), and sleep apnoea (SA). Multiple Sleep Latency Test (MSLT), Maintenance of Wakefulness Test (MWT), Epworth Sleepiness Scale (ESS), Polysomnography (PSG), Actigraphy (ACT), Steer Clear Test (SCT), Psychomotor Vigilance Test (PVT), Pupillary Unrest Index (PUI), and Apnoea-Hypopnoea Index in polysomnography (AHI).

**Table 1: Overview of sleep and vigilance tests and the related parameters relevant to this study**

| Test | Description | Aim | Parameter [unit; range] |
|---|---|---|---|
| Multiple sleep latency test (MSLT) | exposure to a sleep-promoting condition and allowing sleep to occur (4-5 trials spread over one day, determination of the (mean) sleep latency) | assess sleep propensity | sleep latency (SLAT) [min; 0 - 20] |
| Maintenance of wakefulness test (MWT) | exposure to a sleep-promoting condition with the instruction to maintain wakefulness (4-5 trials spread over one day, determination of the (mean) sleep latency) | assess the ability to stay awake | sleep latency (WLAT) [min; 0 - 40] |
| Epworth sleepiness scale (ESS) | 8-item questionnaire (Likert-scale, 0-3 points) to determine the likelihood of falling asleep in specific situations | quantify excessive daytime sleepiness | sleepiness score (ESS) [score; 0 - 24] |
| Steer clear test (SCT) | virtual steering test (possibility to switch between two lanes), instruction not to hit any obstacle ahead representing a 'go/no-go' paradigm | vigilance measurement | error rate (SCTer) [%, 0 - 100] |
| Psychomotor vigilance test (PVT) | visual stimuli prompting a motor reaction | vigilance measurement | reaction time (PVTrt) [ms; ≥ 0] |
| Polysomnography (PSG) | sleep monitoring based on biophysiological signals (e.g. electroencephalogram and breathing) | measurement of all aspects of sleep (e.g. duration or quality) | sleep efficiency (PSE) [fraction; 0 - 1] |
| | | | apnoea-hypopnoea index (AHI) [events/h; ≥ 0] |
| Pupillography (PUP) | pupil recording (i.e. pupillary size and derived functions) in a sleep-promoting environment | measure sleepiness based on autonomous nervous system functions | pupillary unrest index (PUI) [mm/min; ≥ 0] |
| Wrist actigraphy (ACT) | determination of physical activity by sensors on the wrist measuring acceleration, recording of multiple 24h epochs (mostly 1-2 weeks) | assess sleep-wake pattern | inactivity index (ACTI) [%, 0 - 100] |

**Table 2a: Demographics (dataset A)**

| | Overall (%) | NT1 (%) | NT2 (%) | IH (%) | NOH (%) | ISS (%) | FS (%) | SA (%) |
|---|---|---|---|---|---|---|---|---|
| N (%) | 1352 (100) | 106 (7.8) | 90 (6.7) | 119 (8.8) | 192 (14.2) | 205 (15.2) | 183 (13.5) | 457 (33.8) |
| Sex female (%) | 598 (44.2) | 50 (47.1) | 39 (43.3) | 72 (60.5) | 119 (62.0) | 81 (39.5) | 115 (62.8) | 122 (26.7) |
| Age mean [SD] | 40 [16] | 38 [17] | 31 [15] | 27 [10] | 36 [13] | 36 [13] | 37 [14] | 51 [14] |
| **Secondary Diagnosis** | | | | | | | | |
| none | 611 (45.2) | 66 (62.3) | 41 (45.6) | 49 (41.2) | 73 (38) | 66 (32.2) | 106 (57.9) | 210 (46) |
| HYG | 174 (12.9) | - | 4 (4.4) | 9 (7.6) | 23 (12) | 27 (13.2) | 43 (23.5) | 68 (14.9) |
| SA | 117 (8.7) | 23 (21.7) | 14 (15.6) | 6 (5) | 27 (14.1) | 47 (22.9) | - | - |
| SRMD | 115 (8.5) | 10 (9.5) | 8 (8.9) | 3 (2.5) | 8 (4.2) | 9 (4.4) | 9 (4.9) | 68 (14.9) |
| EDSU | 99 (7.3) | - | 9 (10) | 14 (11.8) | 23 (12) | 38 (18.5) | - | 15 (3.3) |
| Insomnia | 57 (4.2) | - | 1 (1.1) | 1 (0.8) | 8 (4.2) | - | 17 (9.3) | 30 (6.6) |
| Parasomnias | 45 (3.3) | 6 (5.6) | 4 (4.4) | 7 (6.3) | 3 (1.5) | 2 (1) | 2 (1.1) | 21 (4.6) |
| FS | 44 (3.3) | - | - | 1 (0.8) | 11 (5.7) | 4 (2) | - | 28 (6.1) |
| ISS | 38 (2.8) | 1 (0.9) | 3 (3.3) | 12 (10.1) | 12 (6.2) | - | 2 (1.1) | 8 (1.8) |
| NOH | 30 (2.2) | - | 5 (5.5) | 17 (14.3) | - | 2 (1) | 2 (1.1) | 4 (0.9) |
| IH | 16 (1.2) | - | 1 (1.1) | - | 3 (1.6) | 5 (2.4) | 2 (1.1) | 5 (1.1) |
| NT2 | 6 (0.4) | - | - | - | 1 (0.5) | 5 (2.4) | - | - |
| **Tertiary Diagnosis** | | | | | | | | |
| none | 1126 (83.3) | 104 (98.1) | 75 (83.7) | 98 (82.4) | 158 (82.3) | 170 (82.9) | 163 (89.1) | 358 (78.3) |
| HYG | 49 (3.6) | - | 3 (3.3) | 8 (6.7) | 8 (4.2) | 14 (6.8) | 4 (2.2) | 12 (2.6) |
| SRMD | 23 (1.7) | - | 2 (2.2) | - | 3 (1.6) | 4 (2) | - | 14 (3.1) |
| Insomnia | 19 (1.4) | - | - | - | 6 (3.1) | 2 (1) | 5 (2.7) | 6 (1.3) |
| EDSU | 14 (1) | - | 3 (3.3) | 4 (3.4) | 1 (0.5) | 3 (1.5) | - | 3 (0.7) |
| Parasomnias | 8 (0.6) | 1 (0.9) | - | 2 (1.7) | 1 (0.5) | 2 (1) | - | 2 (0.4) |
| ISS | 5 (0.4) | - | 3 (3.3) | 1 (0.8) | 1 (0.5) | - | - | - |
| SA | 2 (0.15) | - | 1 (1.1) | - | - | 1 (0.5) | - | - |
| FS | 1 (0.05) | - | - | - | 1 (0.5) | - | - | - |
| **Non-sleep related** | | | | | | | | |
| Parkinson Syndrome | 47 (3.5) | - | - | 1 (0.8) | - | 2 (1) | 1 (0.5) | 43 (9.4) |
| Depression | 41 (3.0) | - | 2 (2.2) | 5 (4.2) | 12 (6.3) | 3 (1.5) | 7 (3.8) | 12 (2.6) |
| Epilepsy | 15 (1.1) | - | 1 (1.1) | - | 1 (0.5) | 4 (2) | 2 (1.1) | 7 (1.5) |
| Multiple Sclerosis | 2 (0.15) | 1 (0.9) | - | - | - | - | 1 (0.5) | - |

**Table 2b: Demographics (dataset B)**

| | Overall (%) | NT1 (%) | NT2 (%) | IH (%) | NOH (%) | ISS (%) | FS (%) | SA (%) |
|---|---|---|---|---|---|---|---|---|
| N (%) | 1101 (100) | 82 (7.5) | 67 (6.0) | 83 (7.5) | 138 (12.5) | 142 (13.0) | 177 (16.0) | 412 (37.5) |
| Sex female (%) | 504 (45.8) | 46 (56.1) | 29 (43.3) | 54 (65.1) | 90 (65.2) | 64 (45.1) | 113 (63.8) | 108 (26.2) |
| Age mean [SD] | 40 [16] | 35 [16] | 30 [14] | 26 [10] | 35 [13] | 34 [14] | 37 [14] | 52 [14] |
| **Secondary Diagnosis** | | | | | | | | |
| none | 611 (55.5) | 66 (80.5) | 41 (61.2) | 49 (59.0) | 73 (52.9) | 66 (46.5) | 106 (59.9) | 210 (51.0) |
| HYG | 174 (15.8) | - | 4 (6.0) | 9 (10.8) | 23 (16.7) | 27 (19.0) | 43 (24.3) | 68 (16.5) |
| SRMD | 115 (10.4) | 10 (12.2) | 8 (12.0) | 3 (3.6) | 8 (5.8) | 9 (6.3) | 9 (5.1) | 68 (16.5) |
| EDSU | 99 (9.0) | - | 9 (13.4) | 14 (16.9) | 23 (16.7) | 38 (26.8) | - | 15 (3.6) |
| Insomnia | 57 (5.2) | - | 1 (1.5) | 1 (1.2) | 8 (5.8) | - | 17 (9.6) | 30 (7.3) |
| Parasomnias | 45 (4.1) | 6 (7.4) | 4 (6.0) | 7 (8.4) | 3 (2.1) | 2 (1.4) | 2 (1.1) | 21 (5.1) |

Abbreviations: Narcolepsy with (NT1) and without cataplexy (NT2), idiopathic hypersomnia (IH), nonorganic hypersomnia (NOH), insufficient sleep syndrome (ISS), fatigue syndromes (FS), sleep apnoea (SA), poor sleep hygiene (HYG), sleep-related movement disorders (SRMD), excessive daytime sleepiness of unknown origin (EDSU).

**Table 3: Descriptive statistics (dataset B)**

| | **Overall** | | | **NT1** | | | **NT2** | | | **IH** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *Parameter [unit]* | *N* | *Median* | *IQR* | *N* | *Median* | *IQR* | *N* | *Median* | *IQR* | *N* | *Median* | *IQR* |
| SLAT [min] | 1101 | 6.9 | 6.5 | 82 | 2.5 | 2.8 | 67 | 3.9 | 3.8 | 83 | 6.8 | 4.8 |
| WLAT [min] | 751 | 29.6 | 23.6 | 38 | 9.9 | 12.1 | 38 | 10.9 | 10.7 | 53 | 27.6 | 20.1 |
| ESS [score] | 1053 | 12 | 7 | 78 | 17 | 5.25 | 63 | 16 | 6 | 82 | 13 | 6 |
| SCTer [%] | 1018 | 4 | 6.8 | 76 | 6.5 | 15.3 | 63 | 5 | 9.6 | 77 | 2 | 5 |
| PVTrt [ms] | 642 | 288 | 107 | 27 | 325 | 578 | 31 | 312 | 347 | 48 | 268 | 79 |
| PSE [fraction] | 1048 | 0.91 | 0.14 | 75 | 0.89 | 0.13 | 60 | 0.93 | 0.09 | 78 | 0.95 | 0.06 |
| PUI [mm/min] | 895 | 7.2 | 5.3 | 58 | 10.3 | 6.9 | 54 | 9.6 | 6.4 | 74 | 9.1 | 6.4 |
| ACTI [%] | 835 | 34 | 8 | 51 | 33 | 6.8 | 51 | 32.9 | 9.3 | 66 | 35 | 8 |
| AHI [events/h] | 1048 | 4.8 | 14.9 | 74 | 1.8 | 4.7 | 60 | 1.5 | 2.7 | 78 | 1.3 | 2 |
| | **NOH** | | | **ISS** | | | **FS** | | | **SA** | | |
| *Parameter [unit]* | *N* | *Median* | *IQR* | *N* | *Median* | *IQR* | *N* | *Median* | *IQR* | *N* | *Median* | *IQR* |
| SLAT [min] | 138 | 6.7 | 4.8 | 142 | 5.7 | 4.6 | 177 | 12.1 | 5.1 | 412 | 6.5 | 5.9 |
| WLAT [min] | 106 | 28.7 | 18.9 | 99 | 32 | 17.25 | 116 | 39.5 | 12.4 | 301 | 30.4 | 23.6 |
| ESS [score] | 136 | 13 | 6 | 138 | 12 | 6.25 | 171 | 10 | 8 | 385 | 12 | 7 |
| SCTer [%] | 129 | 4 | 6 | 130 | 3 | 4.25 | 167 | 3 | 5 | 376 | 4 | 7 |
| PVTrt [ms] | 90 | 308 | 139 | 81 | 270 | 68 | 111 | 271 | 63 | 254 | 302 | 109 |
| PSE [fraction] | 124 | 0.93 | 0.13 | 136 | 0.93 | 0.08 | 164 | 0.89 | 0.14 | 411 | 0.86 | 0.17 |
| PUI [mm/min] | 116 | 7.2 | 4.8 | 117 | 7.2 | 5.2 | 153 | 6 | 4.6 | 323 | 6.5 | 4.6 |
| ACTI [%] | 110 | 34.7 | 6.9 | 112 | 33.5 | 7.2 | 136 | 36.6 | 8.9 | 309 | 33 | 8.6 |
| AHI [events/h] | 125 | 1.5 | 2.8 | 135 | 2.1 | 3.3 | 164 | 2.1 | 3 | 412 | 20 | 23.2 |

Abbreviations: Interquartile range (IQR), Narcolepsy with (NT1) and without cataplexy (NT2), idiopathic hypersomnia (IH), nonorganic hypersomnia (NOH), insufficient sleep syndrome (ISS), fatigue syndromes (FS), sleep apnoea (SA). Sleep latency in Multiple Sleep Latency Test (SLAT), sleep latency in Maintenance of Wakefulness Test (WLAT), Epworth Sleepiness Scale (ESS), sleep efficiency in Polysomnography (PSE), Actigraphy inactivity index (ACTI), Steer Clear Test error rate (SCTer), Psychomotor Vigilance Test reaction time (PVTrt), Pupillary Unrest Index (PUI), Apnoea-Hypopnoea Index in polysomnography (AHI).

**Table 4: Best discriminating SVT and cut-off values (dataset B)**

| Comparison | | 1st | Cut-off | Sens | Spec | 2nd | Cut-off | Sens | Spec | 3rd | Cut-off | Sens | Spec |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **NT1** | NT2 | - | | | | - | | | | - | | | |
| | IH | SLAT | < 4.1 | 0.70 | 0.72 | WLAT | < 19.5 | 0.81 | 0.70 | SCTer | > 3.5 | 0.70 | 0.64 |
| | NOH | SLAT | < 4.6 | 0.89 | 0.79 | WLAT | < 17.6 | 0.81 | 0.78 | ESS | > 15.5 | 0.62 | 0.72 |
| | ISS | WLAT | < 17.6 | 0.81 | 0.79 | SLAT | < 3.6 | 0.76 | 0.79 | ESS | > 14.5 | 0.68 | 0.71 |
| | FS | SLAT | < 6.4 | 0.95 | 0.97 | WLAT | < 22.6 | 0.84 | 0.90 | ESS | > 13.5 | 0.76 | 0.76 |
| | SA | AHI | < 7.6 | 0.87 | 0.95 | SLAT | < 3.9 | 0.70 | 0.81 | ESS | > 13.5 | 0.79 | 0.66 |
| **NT2** | NT1 | - | | | | - | | | | - | | | |
| | IH | WLAT | < 19.5 | 0.81 | 0.70 | SLAT | < 4.1 | 0.70 | 0.72 | SCTer | > 3.5 | 0.70 | 0.64 |
| | NOH | WLAT | < 19.1 | 0.77 | 0.79 | SLAT | < 4.7 | 0.77 | 0.78 | PUI | > 8.8 | 0.71 | 0.69 |
| | ISS | WLAT | < 20.7 | 0.86 | 0.75 | SCTer | > 3.5 | 0.69 | 0.60 | ESS | > 14.5 | 0.60 | 0.70 |
| | FS | SLAT | < 7.2 | 0.94 | 0.94 | WLAT | < 21.2 | 0.89 | 0.92 | ESS | > 11.5 | 0.78 | 0.63 |
| | SA | AHI | < 6.0 | 0.97 | 1.00 | WLAT | < 17.9 | 0.80 | 0.74 | ESS | > 14.5 | 0.60 | 0.73 |
| **IH** | NT1 | SLAT | > 4.1 | 0.72 | 0.70 | WLAT | > 19.5 | 0.70 | 0.81 | SCTer | < 3.5 | 0.64 | 0.70 |
| | NT2 | WLAT | > 19.5 | 0.70 | 0.81 | SLAT | > 4.1 | 0.72 | 0.70 | SCTer | < 3.5 | 0.64 | 0.70 |
| | NOH | PSE | > 0.94 | 0.64 | 0.59 | SCTer | < 3.0 | 0.63 | 0.56 | - | | | |
| | ISS | - | | | | - | | | | - | | | |
| | FS | SLAT | < 8.9 | 0.83 | 0.80 | PSE | > 0.93 | 0.73 | 0.63 | WLAT | < 30.7 | 0.62 | 0.71 |
| | SA | AHI | < 6.7 | 0.96 | 0.98 | PSE | > 0.92 | 0.76 | 0.66 | PUI | > 8.6 | 0.59 | 0.70 |
| **NOH** | NT1 | SLAT | > 4.6 | 0.79 | 0.89 | WLAT | > 17.6 | 0.78 | 0.81 | ESS | < 15.5 | 0.72 | 0.62 |
| | NT2 | WLAT | > 19.1 | 0.79 | 0.77 | SLAT | > 4.7 | 0.78 | 0.77 | PUI | < 8.8 | 0.69 | 0.71 |
| | IH | PSE | < 0.94 | 0.59 | 0.64 | SCTer | > 3.0 | 0.56 | 0.63 | - | | | |
| | ISS | PVTrt | > 284 | 0.63 | 0.65 | SCTer | > 3.5 | 0.57 | 0.69 | - | | | |
| | FS | SLAT | < 9.4 | 0.75 | 0.76 | WLAT | < 32.3 | 0.65 | 0.64 | ESS | > 11.5 | 0.62 | 0.63 |
| | SA | AHI | < 7.1 | 0.96 | 0.98 | PSE | > 0.9 | 0.59 | 0.62 | - | | | |
| **ISS** | NT1 | WLAT | > 17.6 | 0.79 | 0.81 | SLAT | > 3.6 | 0.79 | 0.76 | ESS | < 14.5 | 0.71 | 0.68 |
| | NT2 | WLAT | > 20.7 | 0.75 | 0.86 | SCTer | < 3.5 | 0.60 | 0.69 | ESS | < 14.5 | 0.70 | 0.60 |
| | IH | - | | | | - | | | | - | | | |
| | NOH | PVTrt | < 284 | 0.65 | 0.63 | SCTer | < 3.5 | 0.69 | 0.57 | - | | | |
| | FS | SLAT | < 8.0 | 0.78 | 0.99 | ACTI | < 35.0 | 0.68 | 0.60 | WLAT | < 35.4 | 0.67 | 0.54 |
| | SA | AHI | < 7.0 | 0.96 | 0.97 | PSE | > 0.92 | 0.75 | 0.67 | PVTrt | > 282 | 0.62 | 0.60 |
| **FS** | NT1 | SLAT | > 6.4 | 0.97 | 0.95 | WLAT | > 22.6 | 0.90 | 0.84 | ESS | < 13.5 | 0.76 | 0.76 |
| | NT2 | SLAT | > 7.2 | 0.94 | 0.94 | WLAT | > 21.2 | 0.92 | 0.89 | ESS | < 11.5 | 0.63 | 0.78 |
| | IH | SLAT | > 8.9 | 0.80 | 0.83 | PSE | < 0.93 | 0.63 | 0.73 | WLAT | > 30.7 | 0.71 | 0.62 |
| | NOH | SLAT | > 9.4 | 0.76 | 0.75 | WLAT | > 32.3 | 0.64 | 0.65 | ESS | < 11.5 | 0.63 | 0.62 |
| | ISS | SLAT | > 8.0 | 0.99 | 0.78 | ACTI | > 35.0 | 0.60 | 0.68 | WLAT | > 35.4 | 0.54 | 0.67 |
| | SA | AHI | < 7.3 | 0.93 | 0.96 | SLAT | > 10.0 | 0.73 | 0.74 | ACTI | > 35.1 | 0.61 | 0.63 |
| **SA** | NT1 | AHI | > 7.6 | 0.95 | 0.87 | SLAT | > 3.9 | 0.81 | 0.70 | ESS | < 13.5 | 0.66 | 0.79 |
| | NT2 | AHI | > 6.0 | 0.99 | 0.97 | WLAT | > 17.9 | 0.74 | 0.80 | ESS | < 14.5 | 0.73 | 0.60 |
| | IH | AHI | > 6.7 | 0.98 | 0.96 | PSE | < 0.92 | 0.66 | 0.76 | PUI | < 8.6 | 0.70 | 0.59 |
| | NOH | AHI | > 7.1 | 0.98 | 0.96 | PSE | < 0.9 | 0.62 | 0.59 | - | | | |
| | ISS | AHI | > 7.0 | 0.97 | 0.96 | PSE | < 0.92 | 0.67 | 0.75 | PVTrt | < 282 | 0.60 | 0.62 |
| | FS | AHI | > 7.3 | 0.96 | 0.93 | SLAT | < 10.0 | 0.74 | 0.73 | ACTI | < 35.1 | 0.63 | 0.61 |

Abbreviations and units: Sensitivity (Sens), Specificity (Spec), Narcolepsy with (NT1) and without cataplexy (NT2), idiopathic hypersomnia (IH), nonorganic hypersomnia (NOH), insufficient sleep syndrome (ISS), fatigue syndromes (FS), sleep apnoea (SA). Sleep latency in Multiple Sleep Latency Test [min] (SLAT), sleep latency in Maintenance of Wakefulness Test [min] (WLAT), Epworth Sleepiness Scale [score; 0 – 24] (ESS), sleep efficiency in Polysomnography [fraction; 0 – 1] (PSE), Actigraphy inactivity index [%] (ACTI), Steer Clear Test error rate [%] (SCTer), Psychomotor Vigilance Test reaction time [ms] (PVTrt), Pupillary Unrest Index [mm/min] (PUI), Apnoea-Hypopnoea Index in polysomnography [events/hour] (AHI).

**Table 5: Classification performance (dataset C)**

| | Predicted (SA included) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Observed** | **NT1** | **NT2** | **IH** | **NOH** | **ISS** | **FS** | **SA** | **Correct** |
| NT1 | **44** | 8 | 3 | 7 | 12 | 2 | 6 | 53.7% |
| NT2 | 19 | **17** | 3 | 12 | 11 | 4 | 1 | 25.4% |
| IH | 3 | 6 | **16** | 10 | 28 | 17 | 3 | 19.3% |
| NOH | 6 | 4 | 13 | **42** | 40 | 26 | 7 | 30.4% |
| ISS | 12 | 5 | 10 | 12 | **76** | 19 | 8 | 53.5% |
| FS | 0 | 0 | 6 | 16 | 17 | **126** | 12 | 71.2% |
| SA | 4 | 1 | 0 | 1 | 7 | 9 | **390** | 94.7% |
| Overall | 8.0% | 3.7% | 4.6% | 9.1% | 17.3% | 18.4% | 38.8% | **64.6%** |

| | Predicted (SA excluded) | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Observed** | **NT1** | **NT2** | **IH** | **NOH** | **ISS** | **FS** | **SA** | **Correct** |
| NT1 | **51** | 5 | 6 | 7 | 11 | 2 | - | 62.2% |
| NT2 | 22 | **16** | 2 | 9 | 15 | 3 | - | 23.9% |
| IH | 1 | 7 | **16** | 10 | 31 | 18 | - | 19.3% |
| NOH | 8 | 3 | 12 | **43** | 41 | 31 | - | 31.2% |
| ISS | 12 | 4 | 13 | 12 | **77** | 24 | - | 54.2% |
| FS | 0 | 0 | 4 | 18 | 17 | **138** | - | 78.0% |
| Overall | 13.6% | 5.1% | 7.7% | 14.4% | 27.9% | 31.3% | - | **49.5%** |

Abbreviations: Narcolepsy with (NT1) and without cataplexy (NT2), idiopathic hypersomnia (IH), nonorganic hypersomnia (NOH), insufficient sleep syndrome (ISS), fatigue syndromes (FS), sleep apnoea (SA).

Figure 1

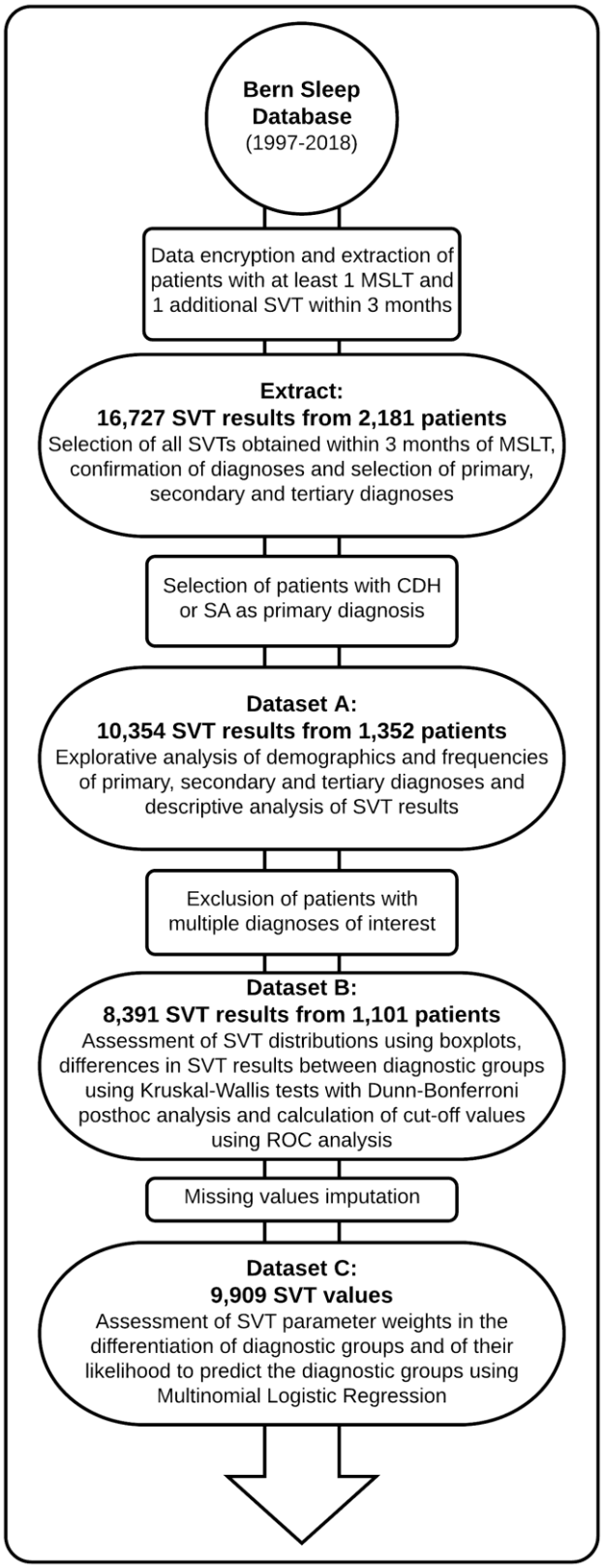

Figure 2

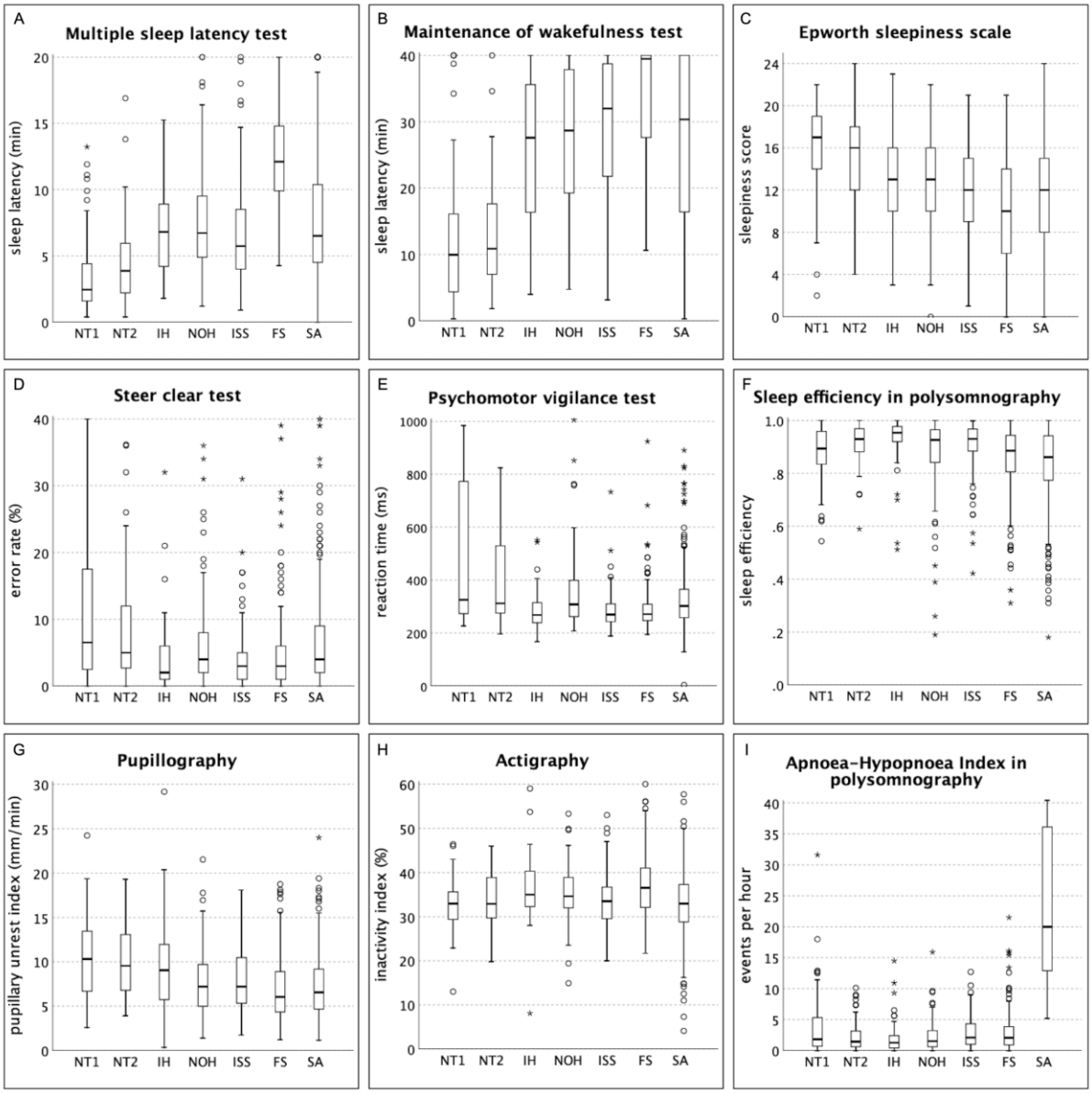

Figure 3