Ullanlinna Narcolepsy Scale in diagnosis of narcolepsy

Sarkanen, Tomi; Alakuijala, Anniina; Partinen, Markku

2019-03
Oxford University Press
http://hdl.handle.net/10138/303482
Sarkanen, T, Alakuijala, A & Partinen, M 2019, 'Ullanlinna Narcolepsy Scale in diagnosis of narcolepsy', Sleep, vol. 42, no. 3, 238. https://doi.org/10.1093/sleep/zsy238

SLEEPJ, 2019, 1–7

doi: 10.1093/sleep/zsy238
Advance Access Publication Date: 26 November 2018
Original Article

ORIGINAL ARTICLE

Ullanlinna Narcolepsy Scale in diagnosis of narcolepsy

Tomi Sarkanen[1,2,*], Anniina Alakuijala[2,3] and Markku Partinen[2,4]

[1]Department of Neurology, Tampere University Hospital, Tampere, Finland, [2]Department of Neurological Sciences, University of Helsinki, Helsinki, Finland, [3]Department of Clinical Neurophysiology, HUS Medical Imaging Center, Helsinki University Central Hospital, Helsinki, Finland and [4]Helsinki Sleep Clinic, Vitalmed Research Centre, Helsinki, Finland

*Corresponding author: Tomi Sarkanen, Department of Neurology, Tampere University Hospital, Z1, 2nd Floor, Teiskontie 35, 33521 Tampere, Finland. Email: tomi.sarkanen@gmail.com.

Abstract

Study objectives: To validate Ullanlinna Narcolepsy Scale (UNS) as a screening tool for narcolepsy in a clinical population and to compare it with Swiss Narcolepsy Scale (SNS) and Epworth Sleepiness Scale (ESS).

Methods: UNS questionnaires of 267 participants visiting Helsinki Sleep Clinic were analyzed. The diagnoses of the participants were narcolepsy type 1 (NT1, $n = 89$), narcolepsy type 2 (NT2, $n = 10$), other hypersomnias ($n = 24$), sleep apnea ($n = 37$), restless legs syndrome or periodic limb movement disorder ($n = 56$), and other sleep-related disorders ($n = 51$). In addition, ESS and SNS scores in a subset of sample (total $N = 167$) were analyzed and compared to UNS.

Results: Mean UNS score in NT1 was 22.0 (95% confidence interval [CI] = 20.4 to 23.6, range 9–43), which was significantly higher than in other disorders, including NT2 (mean 13.7, 95% CI = 10.3 to 17.1, range 7–21, $p = .0013$). Sensitivity and specificity of UNS in separating NT1 from other disorders were 83.5% and 84.1%, respectively. Positive and negative predictive values were 82.5% and 85.1%, respectively. Sensitivities of SNS and ESS in NT1 were 77.2% and 88.6%, and specificities 88.6% and 45.5%, respectively. There were no differences in receiver operating characteristic curves between UNS and SNS. UNS had moderate negative correlation with hypocretin-1 levels ($r_s = -.564$, $p < .001$), and mean sleep latency in multiple sleep latency test ($r_s = -.608$, $p < .001$).

Conclusions: UNS has high specificity and sensitivity for NT1 in a sleep clinic setting. UNS scores below 9 strongly suggest against the diagnosis of narcolepsy.

Statement of Significance

We show that Ullanlinna Narcolepsy Scale has excellent sensitivity and specificity in the screening of narcolepsy type 1 in a clinical population. We also show that if Ullanlinna Narcolepsy Scale score is less than 9, narcolepsy is highly unlikely. Therefore, it might be a feasible approach to interview these patients even more carefully to have better information about other, more common, causes for sleepiness, such as insufficient sleep, circadian rhythm disorder, and sleep apnea. In these patients, sleep diary, actigraphy, and cardiorespiratory polygraphy may be the best next steps in the diagnostic workup instead of full-night polysomnography followed by multiple sleep latency test, as the prior probability of narcolepsy is very low.

Key words: narcolepsy; cataplexy; hypersomnolence; idiopathic; sleep apnea; disorders of excessive somnolence; surveys and questionnaires; Ullanlinna Narcolepsy Scale; Swiss Narcolepsy Scale; hypersomnolence

Submitted: 10 June, 2018; **Revised:** 12 October, 2018

Introduction

Excessive daytime sleepiness (EDS) is a common complaint, even if primary central disorders of hypersomnolence are rare. In addition to clinical history, thorough and expensive diagnostic procedures such as full-night polysomnography, multiple sleep latency test (MSLT), and lumbar puncture are needed for the diagnosis of hypersomnia syndromes. Sleep questionnaires, possibly combined with a sleep log or actigraphy, are cheap and relatively quick tools to help in selecting patients for these intensive diagnostic tests.

Hublin *et al*. introduced the Ullanlinna Narcolepsy Scale (UNS) in 1994 [1]. It was originally designed as a screening tool for narcolepsy. The UNS score varies between 0 and 44, where higher score reflects more symptoms of narcolepsy. The first validation study of UNS showed a sensitivity of 100% and specificity of 98.8% using 14 as a cut point [1]. However, the study was limited in size and used the first version of *International Classification of Sleep Disorders* (*ICSD*) from 1990 [2].

There are only a few other validated narcolepsy screening questionnaires. The Swiss Narcolepsy Scale (SNS) was introduced in 2004 [3]. The SNS score varies between −110 and +66, and scores under 0 indicate possible narcolepsy. SNS has shown sensitivity of 86%–98% and specificity of 86%–96% in separating narcolepsy from different sleep disorders or hypersomnias. Further validation studies of SNS were published as conference posters with limited information on case definition and methods [4, 5], Moreover, a few controversial narcolepsy diagnoses in the first study could have had some impact on the final evaluation of SNS.

In the third version of *International Classification of Sleep Disorders* (*ICSD-3*), narcolepsy is divided into two main categories [6]: narcolepsy type 1 (NT1) and narcolepsy type 2 (NT2). NT1 is a central disorder of hypersomnolence caused by a selective destruction of hypothalamic hypocretin-producing neurons. NT1 is characterized by EDS, disturbed sleep, cataplexy, and parasomnias such as hallucinations and sleep paralyses. The etiology of the less common NT2 is poorly understood as those diagnosed exhibit normal levels of hypocretin-1 (HCRT) in cerebrospinal fluid (CSF). Moreover, cataplexy is absent in NT2. Nonetheless, polysomnography and MSLT criteria (mean sleep latency ≤ 8 min and ≥ 2 sleep onset rapid eye movement sleep periods [SOREMPs]) are similar for NT1 and NT2. Validation studies for UNS or other screening scales have not been conducted according to the *ICSD-3* classification.

Need for a further validation of existing narcolepsy scales was also addressed in a recent review [7]. The authors called for clinically feasible tools that would be easily administered and well validated to distinguish narcolepsy and its subtypes from other hypersomnias. In addition, there is a lack of quantitative means to assess the disease severity.

Our aim was to analyze performance of UNS in the diagnosis of narcolepsy in sleep clinic population and to validate it according to the *ICSD-3*. We also compared performances of UNS, SNS, and Epworth Sleepiness Scale (ESS).

Methods

We reviewed sleep questionnaires of 267 participants at Helsinki Sleep Clinic, Vitalmed Research Center (Helsinki, Finland) between April 15, 2004 and December 2, 2015. The participants were either part of NARPANord narcolepsy project or represented a sample of people with consecutive sleep disorder admitted to full-night polysomnography. We included 89 patients with NT1, 10 with NT2, 37 with obstructive sleep apnea (OSA), 56 with restless legs syndrome or periodic limb movement disorder (RLS/PLMD), 51 with other sleep-related disorders (OSRD), 7 with idiopathic hypersomnia (IH), 3 with Kleine–Levin syndrome (KLS), and 14 with other hypersomnia syndrome (ICD-10 code G47.1). We combined the participants with IH, KLS, and other hypersomnias and defined that entity as "hypersomnia" (HS).

Only participants with complete UNS data were included in all the analyses. In sensitivity, specificity, and predictive value analyses with comparisons to SNS and ESS, only a dataset without any missing values in SNS or ESS was applied. In these data, there were 79 NT1, 9 NT2, 22 OSA, 12 RLS/PLMD, 32 OSRD, and 13 HS participants.

The diagnoses were made according to the *ICSD-2* or the *ICSD-3* [6]. NT1 participants had EDS and either low HCRT levels in the CSF sample or, if a lumbar puncture had not been made, both clear-cut cataplexy and positive MSLT. In addition, they were positive for HLA-DQB1*06:02. NT2 participants had EDS and either normal hypocretin (measured in 8 of 10) or no cataplexy, and positive MSLT. A case was classified as HS if NT1 and NT2 were excluded, and if he/she had severe daytime somnolence without other explaining factors (such as circadian rhythm disorder, insufficient sleep, or sleep-disordered breathing [SDB]). The OSA group was diagnosed with mild to severe sleep apnea. The OSRD group consisted of insomnia (n = 16), delayed sleep phase syndrome (n = 11), depression (n = 10), attention deficit hyperactivity syndrome (n = 4), behaviorally induced insufficient sleep (BIIS) (n = 3), fatigue (n = 3), parasomnias (n = 2), pediatric autoimmune neuropsychiatric syndrome (n = 1), and REM sleep behavior disorder (n = 1). In addition, sleep questionnaires of 85 relatives of narcoleptic participants were analyzed. Relatives were siblings or parents of NT1 participants and were not diagnosed with OSRD, although few of them reported mild and occasional parasomnias or some symptoms of RLS in sleep questionnaires.

Ullanlinna Narcolepsy Scale

UNS consists of 11 items that are graded from 0 to 4 (never to daily/almost daily or > 40 min to <10 min): frequency of cataplexy (4 items), daytime napping (1 item), unintentional sleep lapses (5 items), and sleep latency (1 item), yielding a total score of 0–44 (Supplementary Material 1). The first validation study population included only 24 patients with narcolepsy who had been diagnosed according to *ICSD-1* (Supplementary Material 2) [1].

Swiss Narcolepsy Scale

SNS consists of five questions (Q1–Q5) that are rated on a 5-point scale from 1 (never) to 5 (almost always) [3]. Scores are then combined using an equation: 6 × Q1 + 9 × Q2 − 5 × Q3 − 11 × Q4 − 13 × Q5 + 20 [3]. For our study, SNS was calculated with some adjustments from the Basic Nordic Sleep Questionnaire (BNSQ), UNS, and 5-item World Health Organization Well-Being Scale (WHO-5) questions (Supplementary Material 1) [1, 8, 9].

Statistical analyses

All statistical analyses were conducted using Stata/SE 14.1 for Mac (StataCorp. 2015; College Station, TX: StataCorp LP). Between-group differences were evaluated using analysis of variance, *t*-tests and χ^2 tests as appropriate. Area under curve comparisons were performed with roccomp command in Stata/SE 14.1. Spearman's rank correlation was used to assess relationship between UNS and CSF HCRT levels, because the distribution of the latter was highly skewed to the left. All *p* values were two sided and the significance level was set at 0.05.

Results

The characteristics of our study population are presented in Table 1. The mean age in NT1 group was 23.7 years (*SD* 11.9). They were statistically significantly younger than participants in other groups, except for NT2 group where the mean age was 23.6 years (*SD* 8.1), $p = .583$. All NT1 participants were positive for HLA-DQB1*06:02 allele (one missing). CSF HCRT was below 110 pg/mL in 65 of the 69 tested NT1 participants. The remaining four all had their CSF HCRT levels under 150 pg/mL. Sixty-four percent patients with NT1 had narcolepsy associated with Pandemrix vaccination.

Seventy-four percent patients with NT1 had two or more SOREMPs in MSLT (Table 1). Data on MSLT were missing for 15 patients who had their MSLT done elsewhere. All these patients had low HCRT levels (below 110 pg/mL). Eighty-nine percent patients with NT1 and all patients with NT2 had a mean sleep latency of less than 8 min. Twenty-two to thirty-three percent patients with OSA, HS, or OSRD also had two or more SOREMPs in MSLT, and three of them even fulfilled MSLT criteria for narcolepsy. Diagnoses of these three patients were KLS, severe delayed sleep phase syndrome, and severe OSA.

The results of UNS were similar in all age groups (Figure 1). Mean (*M*) UNS score was higher in NT1 group than in any other group (*M* 22.0, 95% confidence interval [CI] = 20.4 to 23.6, $p < .001$ in all comparisons). Interestingly, in NT2 group, mean UNS score was below 14 and was also statistically significantly lower than in NT1 group (*M* 13.7, 95% CI = 10.3 to 17.1, $p = .0013$). The lowest UNS score in NT1 group was 9. When the questions regarding cataplexy were removed, the mean score in NT2 decreased only slightly, and the difference between NT1 and NT2 disappeared (*M* 14.6, 95% CI = 13.6 to 15.6 vs. *M* 12.8, 95% CI = 9.0 to 16.6, $p = .269$).

Similar findings were seen in SNS. Mean SNS score in NT1 was –32.4 (95% CI = –40.2 to –24.6). Again, SNS scores were low compared to any other group. In NT2, SNS scores were mostly above zero (*M* 16.0, 95% CI = –4.0 to 36.0).

By contrast, ESS total points did not separate NT1 and NT2 from each other but were higher in both narcolepsy syndromes than in HS, OSA, or any other group (Table 1).

There were no statistically significant differences in UNS scores between patients with H1N1–vaccine related and sporadic narcolepsy.

Table 1. Characteristics of study population

	NT1 (*n* = 89)	NT2 (*n* = 10)	HS (*n* = 24)	OSA (*n* = 37)	OSRD (*n* = 51)	RLS/PLMD (*n* = 56)	Relatives (*n* = 85)
Mean age, yrs	23.7	23.6	35.0	52.7	32.1	54.5	39.1
(*SD*)	(11.9)	(8.1)	(15.9)*	(17.3)*	(13.8)*	(15.3)*	(17.8)*
<18, *n*	38	2	2	3	9	0	14
18–40, *n*	41	8	13	3	25	10	25
>40, *n*	10	0	9	31	17	46	46
Mean age at onset, yrs	18.7	21.6	15.4	UK	18.4	UK	N/A
(*SD*)	(10.6)	(7.7)	(5.5)		(8.6)		
Mean UNS	22.0	13.7*	9.7*	6.9*	7.2*	6.0*	4.3*
(*SD*)	(7.6)	(4.8)	(6.5)	(5.7)	(5.4)	(3.4)	(2.4)
Min, max	9, 43	7, 21	1, 30	0, 27	0, 25	1, 17	0, 10
Mean ESS	16.2	14.0	12.2*	7.8*	9.2*	8.4*	4.3*
(*SD*)	(4.7)	(6.1)	(5.3)	(6.1)	(6.3)	(5.4)	(3.1)
Min, max	0, 24	2, 20	0, 20	0, 20	0, 21	1, 17	0, 12
Mean SNS	–32.4	16.0*	23.2*	38.1*	28.5*	26.1*	N/A
(*SD*)	(35.6)	(26.1)	(14.8)	(16.8)	(26.3)	(16.9)	
Min, max	–110, 45	–18, 48	–12, 42	6, 61	–51, 66	1, 56	
HLA-DQB1*06:02 positive *n*/*n* studied	88/88	3/7	7/17	1/8	9/20	1/6	53/79
(% of *n* studied)	(100)	(43)*	(41)*	(13)*	(45)*	(17)*	(67)*
Mean hypocretin-1, pg/mL	36.7	248.6*	272.5*	320.5*	240.6*	297.3*	N/A
(*SD*)	(38.2)	(58.0)	(57.6)	(38.9)	(56.1)	(61.1)	
Hypocretin-1 < 110, pg/mL *n*/*n* studied	65/69†	0/8	0/11	0/2	0/12	0/3	N/A
Mean SL in MSLT SL, min (*SD*)	3.7	4.1	8.7	9.8	13.1	12.5	N/A
	(3.5)	(2.3)	(5.6)	(3.7)	(3.5)	(5.5)	
Mean number of SOREMPs in MSLT	2.7	3.4	0.4	1.2	1.1	0	N/A
(*SD*)	(1.7)	(1.3)	(0.8)	(1.5)	(1.7)	(0)	
Occurrence of MSLT SOREMP ≥ 2 *n*/*n* studied	55/74	9/9	2/19	2/6	7/28	0/9	N/A
(% of *n* studied)	(74)	(100)	(22)	(33)	(25)	(0)	

SD, standard deviation; SL, mean sleep latency; N/A, not applicable; UK, unknown.

*$p < .05$ compared to NT1.

†All four had values between 110 and 150 pg/mL.

If only the dataset of participants without any missing values in SNS and ESS was used (N = 167), sensitivity of UNS at least 14 against other disorders was 83.5%, and specificity 84.1%, positive predictive value (PPV) 82.5%, and negative predictive value (NPV) 85.1% (Table 2).

Sensitivity and specificity of SNS in differentiating NT1 from other disorders were 77.2% and 88.6% (PPV 85.9%, NPV 81.3%), respectively, and 72.7% and 91.1% (PPV 90.1%, NPV 75.0%) in separating both NT1 and NT2 from others (Table 2). Accordingly, if UNS cut point is raised to 17, it provides similar, somewhat lower sensitivity (76.0%) than cut point of 14, but the specificity increases to 90.9%. In confirmed hypocretin deficiency, the sensitivity of SNS in differentiating NT1 from other disorders was 71.4%, specificity 88.6%, PPV 81.8%, and NPV 81.3%.

Sensitivity and specificity of ESS (≥11 points) in NT1 were 88.6% and 45.5%, respectively (PPV 59.3%, NPV 81.6%). If the two narcolepsy syndromes were combined, the sensitivity and specificity were 87.5% and 48.1%, respectively (PPV 65.3%, NPV 77.6%).

The area under a receiver operating characteristic curve (ROC AUC) was large in both UNS and SNS. There were no statistically significant differences between UNS (ROC AUC = .928, 95% CI = 0.891 to 0.963) and SNS (ROC AUC = .921, 95% CI = 0.887 to 0.964), X^2 (1, N = 167) is equal to .01, p value of .921. The ROC AUC was significantly smaller in ESS (ROC AUC = .784, 95% CI = 0.704 to 0.843) than in UNS (X^2 = 25.58, $p < .001$) or in SNS (X^2 = 14.54, $p < .001$; Figure 2). Only patients with results from all questionnaires were included in the ROC AUC analysis.

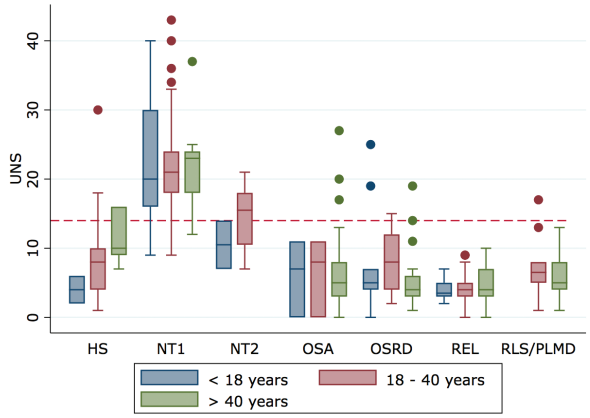

Figure 1. Ullanlinna Narcolepsy Scale by age groups (<18, 18–40, >40 years). Dashed line indicates Ullanlinna Narcolepsy Scale score of 14. NT1, narcolepsy type 1; NT2, narcolepsy type 2; HS, hypersomnia; OSA, obstructive sleep apnea; OSRD, other sleep-related disorders; REL, relatives; RLS, restless legs or periodic limb movement disorder.

Using the whole dataset (N = 267), sensitivity and specificity of UNS at least 14 in differentiating NT1 from NT2, HS, OSA, OSRD, and RLS/PLMD combined were 85.4% and 87.6%, respectively (Table 3). PPV of UNS at least 14 for NT1 was 77.6% and NPV 92.3%. With 13 as a cut point, PPV was 73.2% and NPV 93.7%. If the two narcolepsy syndromes were combined and compared against other disorders, the sensitivity was 82.8% and specificity 90.5%. Scores above 30 gave 100% specificity for NT1 (Table 3). If only patients with NT1 with confirmed hypocretin deficiency were included in the analysis (n = 69), the sensitivity of UNS for NT1 was 84.1%, specificity 87.6%, PPV 72.5%, and NPV 93.4%.

There was a negative correlation between UNS and CSF HCRT levels (r_s = –.564, $p < .001$; Figure 3). There was also a similar correlation between UNS and mean sleep latency in MSLT (r_s = –.608, $p < .001$).

Twenty-three NT1 participants did not have any medication at the time of questionnaires. There were no statistically significant differences in mean UNS scores between medicated and unmedicated participants (M 22.2, 95% CI = 20.5 to 24.0 vs. M 21.8, 95% CI = 17.9 to 25.7, $p = .083$).

Cronbach's alpha for the UNS was 0.898, indicating a high internal consistency of the scale.

Discussion

Our results show that the UNS can be used as a screening tool for narcolepsy in a clinical population. UNS is especially useful in recognizing NT1 and demonstrates a sensitivity of 84%–85% and specificity of 84%–88% against other sleep disorders including NT2 and other hypersomnias. If the two narcolepsy syndromes are combined, UNS still works rather well as sensitivity and specificity remain 80% and 88%, respectively. Differentiating HS from narcolepsy syndromes by UNS is more challenging as the

Table 2. Sensitivity and specificity of cut points of UNS, SNS, and ESS in separating different disorders. Data presented as percentages (%). Participants with missing data in SNS and ESS omitted

	UNS ≥ 13	UNS ≥ 14	UNS ≥ 17	SNS < 0	ESS ≥ 11	
	Sensitivity	Sensitivity	Sensitivity	Sensitivity	Sensitivity	*n*
NT1 vs.	87.3	83.5	76.0	77.2	88.6	79
	Specificity	Specificity	Specificity	Specificity	Specificity	*n*
OSA	86.4	95.5	95.5	100.0	63.6	22
RLS/PLMD	66.7	91.7	91.7	100.0	33.3	12
OSRD	90.6	90.6	93.8	84.4	50.0	32
HS	61.5	69.2	92.9	84.6	30.8	13
NT2	44.4	44.4	66.7	66.7	22.2	9
ALL + NT2	77.3	84.1	90.9	88.6	45.5	88
	Sensitivity	Sensitivity	Sensitivity	Sensitivity	Sensitivity	*n*
NT1 and NT2 vs.	84.1	80.7	71.6	72.7	87.5	88
	Specificity	Specificity	Specificity	Specificity	Specificity	*n*
HS	61.5	69.1		84.6	30.8	13
ALL	81.0	88.6	93.7	91.1	48.1	79

N/A, not applicable. ALL includes HS, OSA, OSRD, RLS/PLMD.

specificity is reduced to 69%. Although the mean UNS scores in NT1 are higher than in NT2, a cut point of 14 does not differentiate these syndromes reliably enough from each other (specificity 44%). Owing to sleepiness, patients with NT2 have UNS scores near 14, but the lack of cataplexy lowers the points below the cut point. Conversely, NT1 participants without any cataplexy had mean UNS scores above 14. However, there was no statistically significant difference between NT1 without cataplexy and NT2, implicating that sleepiness measured by UNS is similar in both diseases. However, the sample size in this comparison is rather small to make strong conclusions on this matter.

UNS scores were 9 or higher in all patients with NT1. Considering our large total sample size (N = 267), we can state that if UNS is below 9, the diagnosis of NT1 is unlikely. This finding could be valuable in the diagnostic workup. For instance, in a case of a sleepy patient with UNS of 8 points, the underlying cause for sleepiness is most probably something else than narcolepsy. Therefore, in such a case it might be a practical approach to screen for SDB or PLMD by cardiorespiratory polygraphy first. Complementary laboratory diagnostics to rule out other somatic diseases may also be warranted, instead of admitting the patient directly to full-night polysomnography followed by MSLT.

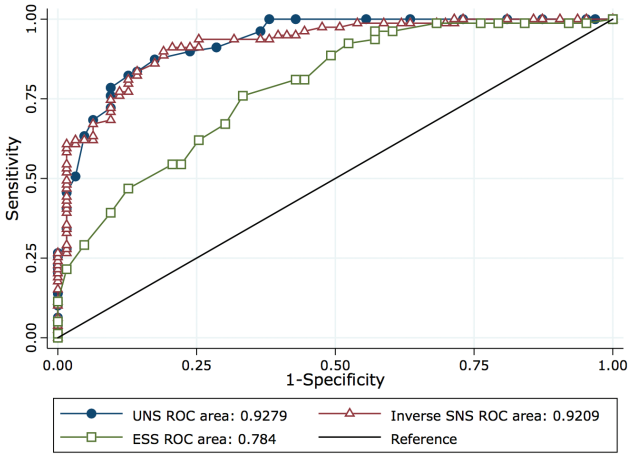

Figure 2. Performance of UNS, SNS, and ESS in differentiating narcolepsy type 1 against narcolepsy type 2, other hypersomnias, restless legs and periodic limb movement disorder, obstructive sleep apnea, and other sleep-related disorders combined. With SNS, opposite values (e.g., 30 to –30) were used in comparison. ROC, receiver operating characteristic. UNS, Ullanlinna Narcolepsy Scale; SNS, Swiss Narcolepsy Scale; ESS, Epworth Sleepiness Scale; ROC, receiver operating characteristic.

Specificity of UNS increases with the cut point. All participants with 31 or more points had NT1 (specificity 100%). A cut point of 17 achieved a specificity more than 90%—the diagnosis of narcolepsy is very likely. As expected, specificity increased if the cut point is elevated (Table 2). Still, 14 as a cut point shows a nice compromise with relatively few false negatives or false positives. UNS cut point values 13 and 14 show almost equal positive likelihood ratios, both higher than 5. A cut point of 13 has a good sensitivity without a marked loss in specificity. If validated in further studies, 13 as a cut point might be more feasible, especially in samples with a low pretest likelihood of narcolepsy. All in all, UNS catches NT1 quite well, and somewhat separates it from NT2 and HS, but at the same time, leaves these two syndromes unseparated, which can be either an advantage or a disadvantage depending on the setting.

Performances of UNS and SNS are quite similar. ROC areas did not differ between the two scales and SNS demonstrated an adequate sensitivity of predicting NT1, even though in our study it had slightly lower sensitivity for NT1 than UNS (77.2% vs. 83.5%). ESS cannot be used as a screening tool for hypersomnia syndromes due to its low specificity and low PPV. Poor performance of ESS was seen also in ROC area, which was significantly lower than in UNS or SNS (Figure 2). Moreover, we saw a moderate correlation between UNS and CSF hypocretin levels and between UNS and sleep latency in MSLT. Chervin *et al.* have previously reported that ESS has only moderate correlation with MSLT (Spearman's r_s = –.37, p = .004) [10]. In our study, the correlation was even lower (r_s = –.316, p < .001). Moreover, another article even found no correlation between ESS and mean sleep latency or severity of OSA [11].

The moderate negative correlation between UNS and HCRT levels is displayed in Figure 3. False negatives are displayed in the lower left rectangle and false positives in higher right part of the figure separated by a dotted line. It is noteworthy that the lumbar puncture was only made if narcolepsy was suspected.

Table 3. Detailed report of sensitivity and specificity of UNS with different cut points in NT1 (n = 89) versus NT2, HS, RLS/PLMD, OSA, and OSRD (n = 178) combined

Cut point	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)	Correctly classified (%)	LR+	LR–
≥9	**100.0**	**69.1**	**61.8**	**100.0**	**79.4**	**3.24**	**0.00**
≥10	96.6	73.6	64.7	97.8	81.3	3.66	0.05
≥11	92.1	78.7	68.3	95.2	83.2	3.47	0.10
≥12	91.0	80.9	70.4	94.7	84.3	4.32	0.11
≥13	**88.8**	**83.7**	**73.2**	**93.7**	**85.4**	**5.45**	**0.13**
≥14	**85.4**	**87.6**	**77.6**	**92.3**	**86.9**	**6.91**	**0.17**
≥15	**84.3**	**88.8**	**79.0**	**91.9**	**87.3**	**7.50**	**0.18**
≥16	80.9	91.0	81.8	90.5	87.6	9.00	0.21
≥17	78.7	93.3	85.4	89.7	88.4	11.67	0.23
≥18	74.2	94.4	86.8	88.0	87.6	13.20	0.27
≥19	70.8	96.1	90.0	86.8	87.6	18.00	0.30
≥20	64.0	97.2	91.9	84.4	86.1	22.80	0.37
—	—	—	—	—	—	—	—
≥31	**13.5**	**100.0**	**100.0**	**69.8**	**71.2**	**N/A**	**0.86**

The bolded values note cut points of special interest.
N/A, not applicable; LR+, positive likelihood ratio; LR–, negative likelihood ratio; LR+ at least 5 is considered moderate increase in the likelihood of disease.

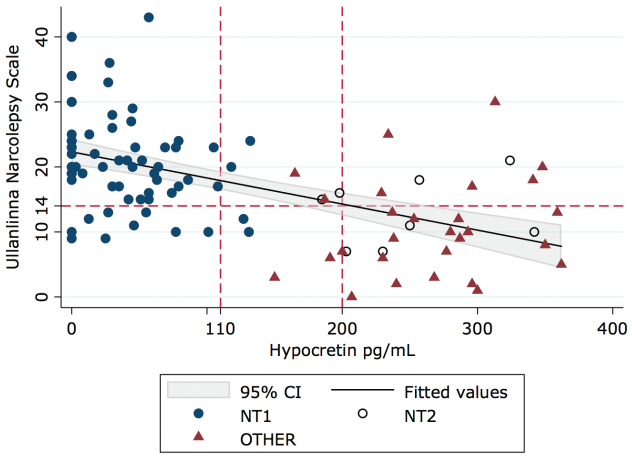

Figure 3. Correlation between Ullanlinna Narcolepsy Scale and hypocretin-1 levels. NT1, narcolepsy type 1; NT2, narcolepsy type 2; OTHER, other sleep disorders.

Therefore, the figure includes a selection bias and does not fully represent the whole-study sample. If all the study participants would have been tested for CSF HCRT, presumably more cases would have been added to the lower right part of the figure. In other words, if we had more test results, we could perhaps show even stronger negative correlation between UNS and HCRT levels. The middle rectangles in the figure are intriguing. These participants have HCRT levels of 110–200 pg/mL, which is considered a gray area. UNS scores also scatter almost equally in this area.

As a matter of fact, UNS has a higher sensitivity for NT1 and HCRT deficiency than the presence of two or more SOREMPs in MSLT, which was seen in only 74% of our NT1 participants. Conversely, 22%–33% participants with HS, OSA, or OSRD who underwent MSLT had at least two SOREMPs, which could lead to wrong narcolepsy diagnoses, especially in the HS group. These findings are in line with other published reports [12–14]. The specificity of MSLT could be increased by reducing mean sleep latency limit to 5 min and analyzing both the proportion of REM sleep in all the naps and the sequence of occurring sleep stages—whether REM sleep occurs before stage N2 sleep—but this would also reduce the sensitivity to around 50% [15, 16]. A test–retest reliability of MSLT is also limited, especially in other diseases than NT1 [17]. UNS is not supposed to replace MSLT in the diagnosis as they are two totally different tests. Proper use of UNS could include using it to asses a priori probability of positive MSLT. In a case of very low UNS points (<9), other diagnostic procedures than MSLT, as demonstrated previously in discussion, might provide the cause for EDS earlier. Furthermore, if MSLT is negative, but the participants had high UNS scores, retesting or CSF hypocretin measurement could be considered.

Number of factors could explain the reasons for the unnoticeable effects of medication on UNS scores. Psychometric properties of UNS in measuring change over time have not been studied. Thus, we do not know how consistent UNS scores are between two visits without a change in health or symptoms of narcolepsy (test–retest reliability). We saw in our previous longitudinal study that there are great individual differences and heterogeneity in the evolution of UNS scores between individuals [18]. The result could also mirror heterogeneity of the disease itself. Some patients having narcolepsy with similar sleep study results or HCRT levels have large differences in the health-related quality of life. Some patients might be able to manage their daily activities without medication whereas others are severely disabled. Reasons for this heterogeneity are unknown. Changes in other neurotransmitter systems (e.g. histamine) could possibly compensate for the loss of hypocretin [19]. Psychological factors such as differences in resilience and comorbidities, e.g., depression, could also affect how symptoms of narcolepsy are experienced [20, 21].

Strengths of our study include large sample of NT1 participants. The vast majority of the participants had HCRT values measured, which eliminates the possibility for wrong positive NT1 diagnoses. It remains an open question if UNS could be used as a measure of severity of narcolepsy, but moderate negative correlation with both HCRT levels and mean sleep latency in MSLT indicate that it could also be suitable for this purpose. Further studies are needed to validate UNS in other settings.

Our study does have some noteworthy limitations. First, due to the retrospective nature of the study, SNS was calculated from the questions in the BNSQ (SNS has no official Finnish translation). Even though these questions are very similar to SNS questions, inconsistency between the original version and our questions is possible. Unfortunately, we had missing data on SNS but to avoid possible biases we used only data without any missing values to compare different scales. Nevertheless, our main aim was to demonstrate the effectiveness of UNS, not explicitly in comparison to SNS. Second, we used a Finnish version of UNS, which could reduce the generalization of our results to other populations. The nomenclature of UNS was not revised for this study and it is possible, at least in theory, that the original English version of UNS might not accurately translate to other languages. However, the wording in UNS, e.g., in question "When laughing, becoming glad or angry, or in an exciting situation, have the following symptoms suddenly occurred" is rather close to the wording on similar questions in Narcolepsy severity scale ("*… when experiencing emotions (laughter, intense pleasure, surprise)*") and in SNS ("*… during emotions like laughing, happiness, and anger*") [22]. Third, this study was conducted in a single sleep clinic setting that can offer a biased population set. In particular, the patients with narcolepsy in our clinic, which acts also a national tertiary center for narcolepsy, may overly represent diagnostically challenging forms of narcolepsy.

Conclusions: Proposition for Clinical Use of UNS

Obviously, sleep questionnaires cannot replace a proper clinical interview and examination in the diagnosis of hypersomnia syndrome. Nonetheless, they offer a valuable tool and aid for clinical practice in this task that can sometimes be very challenging. Our study suggests that one application of these findings could be admitting participants with UNS at least 13 or 14 directly to full-night polysomnography and MSLT. Conversely, if a participant has positive MSLT for narcolepsy (mean sleep latency ≤ 8 min and 2 ≥ SOREMPs) but UNS less than 9, MSLT findings should be interpreted with caution along with any prospective diagnosis of narcolepsy. In such a case, other factors that may cause false positives, such as BIIS, circadian rhythm sleep–wake disorders, SDB, and other sleep disorders,

need to be carefully excluded. On the other hand, UNS at least 31 with normal MSLT may suggest an incorrect negative MSLT. A following step to diagnose/exclude NT1 could include measuring CSF HCRT levels.

Supplementary Material

Supplementary material is available at *SLEEP* online.

Acknowledgments

We thank Anne Huutoniemi for helping in the collection of the samples and Eric Burns for linguistic advice.

Funding

This study was supported by the Academy of Finland (grant no. 260603, NARPANord) and grant to T.S. from the Finnish Medical Foundation.
Dr. Sarkanen reports grant from The Finnish Medical Foundation. Dr. Alakuijala and Dr. Partinen have nothing to disclose.
Conflict of interest statement

References

1. Hublin C, *et al*. The Ullanlinna Narcolepsy Scale: validation of a measure of symptoms in the narcoleptic syndrome. *J Sleep Res*. 1994;**3**(1):52–59.
2. Thorpy M.; Diagnostic Classification Steering Committee. *International Classification of Sleep Disorders: Diagnostic and Coding Manual*. Rochester: American Sleep Disorders Association; 1990.
3. Sturzenegger C, *et al*. The clinical spectrum of narcolepsy with cataplexy: a reappraisal. *J Sleep Res*. 2004;**13**(4): 395–406.
4. American Academy of Sleep Medicine. *International Classification of Sleep Disorders*. 3rd ed. Darien, IL: American Academy of Sleep Medicine; 2014.
5. Kallweit U, *et al*. Patient-reported measures of narcolepsy: the need for better assessment. *JCSM*. 2017;**13**(5): 737–744.
6. Partinen M, *et al*. Basic Nordic Sleep Questionnaire (BNSQ): a quantitated measure of subjective sleep complaints. *J Sleep Res*. 1995;**4**(S1):150–155.
7. Topp CW, *et al*. The WHO-5 well-being index: a systematic review of the literature. *Psychother Psychosom*. 2015;**84**(3):167–176.
8. Sturzenegger C, *et al*. Swiss Narcolepsy Scale: a valid tool for the identification of hypocretin-1 deficient patients with narcolepsy. *J Sleep Res*. 2014;**23** (Suppl. 1):297.
9. Sturzenegger C, *et al*. Swiss Narcolepsy Scale: a valid tool for the identification of hypocretin-1 deficient patients with narcolepsy. *J Sleep Res*. 2006;**15** (Suppl. 1):74.
10. Chervin RD, *et al*. Comparison of the results of the Epworth sleepiness scale and the multiple sleep latency test. *J Psychosom Res*. 1997;**42**(2):145–155.
11. Chervin RD, *et al*. The Epworth sleepiness scale may not reflect objective measures of sleepiness or sleep apnea. *Neurology*. 1999;**52**(1):125–131.
12. Kretzschmar U, *et al*. Which diagnostic findings in disorders with excessive daytime sleepiness are really helpful? A retrospective study. *J Sleep Res*. 2016;**25**(3):307–313.
13. Johns MW. Sensitivity and specificity of the multiple sleep latency test (MSLT), the maintenance of wakefulness test and the Epworth sleepiness scale: failure of the MSLT as a gold standard. *J Sleep Res*. 2000;**9**(1):5–11.
14. Aldrich MS, *et al*. Value of the multiple sleep latency test (MSLT) for the diagnosis of narcolepsy. *Sleep*. 1997;**20**(8):620–629.
15. Murer T, *et al*. Optimizing MSLT specificity in narcolepsy with cataplexy. *Sleep*. 2017;**40**(12). zsx173.
16. Marti I, *et al*. Multiple sleep latency measures in narcolepsy and behaviourally induced insufficient sleep syndrome. *Sleep Med*. 2009;**10**(10):1146–1150.
17. Lopez R, *et al*. Test–retest reliability of the multiple sleep latency test in central disorders of hypersomnolence. *Sleep*. 2017;**40**(12). zsx164.
18. Sarkanen T, *et al*. Clinical course of H1N1-vaccine-related narcolepsy. *Sleep Med*. 2016;**19**:17–22.
19. John J, *et al*. Greatly increased numbers of histamine cells in human narcolepsy with cataplexy. *Ann Neurol*. 2013;**74**(6):786–793.
20. Szakács A, *et al*. Psychiatric comorbidity and cognitive profile in children with narcolepsy with or without association to the H1N1 influenza vaccination. *Sleep*. 2015;**38**(4): 615–621.
21. Bruck D. The impact of narcolepsy on psychological health and role behaviours: negative effects and comparisons with other illness groups. *Sleep Med*. 2001;**2**(5):437–446.
22. Dauvilliers Y, *et al*. Measurement of narcolepsy symptoms: the narcolepsy severity scale. *Neurology*. 2017;**88**(14):1358–1365.